Marine Micropaleontology 174 (2022) 102094

Contents lists available at ScienceDirect

Marine Micropaleontology

journal homepage: www.elsevier.com/locate/marmicro

Ostracods in databases: State of the art, mobilization and future applications

Huai-Hsuan M. Huang [a,b,*], Moriaki Yasuhara [b,*], David J. Horne [c], Vincent Perrier [d], Alison J. Smith [e], Simone N. Brandão [f]

[a] *Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington DC 20013-7012, United States of America*
[b] *School of Biological Sciences, Area of Ecology & Biodiversity, Swire Institute of Marine Science, State Key Laboratory of Marine Pollution, University of Hong Kong, Kadoorie Biological Sciences Building, Pokfulam Road, Hong Kong Special Administrative Region*
[c] *School of Geography, Queen Mary University of London, Mile End Road, London E1 4NS, UK*
[d] *Université de Lyon, UCBL, ENSL, CNRS, UMR 5276 LGL-TPE, 69622 Villeurbanne, France*
[e] *Department of Geology, Kent State University, Kent, OH 44242, USA*
[f] *State University of Santa Cruz (Brazil), Graduate Program in Zoology, Biological Sciences Department, Campus Soane Nazaré de Andrade, Rodovia Jorge Amado, km 16, Bairro Salobrinho, Ilhéus, Bahia, CEP 45662-900, Brazil*

A R T I C L E I N F O

Keywords:
Biogeography
Macroecology
Collaboration tool
Open science
Biodiversity informatics

A B S T R A C T

Biodiversity databases are changing the longevity of data in the era of open science. They also represent a collaboration opportunity in analyzing large-scale (paleo)biological patterns beyond a local project or a time scale. Ostracods, microscopic crustaceans, are a component in many biodiversity databases. They live in most kinds of aquatic environments today and their fossil record spans nearly the whole of the Earth's metazoan biosphere history from Ordovician to Holocene. Thus, ostracods provide an ideal model system for understanding large-scale biodiversity patterns and dynamics in both space and time. Thanks to many contributors, current and future ostracodologists have access to databases that have gone through numerous improvements and have been populated by many datasets. However, rapid growth of databases has caused confusion among users regarding available data, technical terms and database aims. We review key databases that include ostracods, summarizing their history of development, current spatial and temporal coverage, various types of data models and the intertwined relationships between databases. We also present a quantitative summary of ostracod diversity history based on the Paleobiology Database. Our investigations show that the database field is transitioning from the traditional single focus to multipurpose, from static to dynamic data display/download and from independent systems to collaborative networks. We compare the ways several databases approach persistent challenges such as taxonomic harmonization, validation of the original sampling metadata and paleolocality uncertainties. With increasing capability of data integration, databases continue to require enormous efforts regarding high-quality data entry and careful coordination among scientists and technical teams.

1. Introduction

To assess the biological impacts of global changes there is an increasing demand to understand the evolving dynamics among ecological components over various spatial and temporal scales (e.g., Jablonski and Sepkoski Jr, 1996; Scholes et al., 2008; Rick and Lockwood, 2013). Playing an increasingly important role in this research trend, many databases have been developed to preserve, curate, and mobilize hard-earned biodiversity data (Goddard et al., 2011), motivated by a need to increase the discoverability or accessibility of such data. Investigations into macroecology, ecosystem dynamics, and macroevolution increasingly rely on data drawn from biodiversity databases (e.g., Liow and Stenseth, 2007; Marx and Uhen, 2010; Tittensor et al., 2010; Villéger et al., 2011; Lazarus et al., 2014; Kocsis et al., 2018b; Reddin et al., 2020; Chaudhary et al., 2021). The importance of biodiversity databases is evident, and their continuous improvement

* Corresponding authors at: Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington DC 20013-7012, United States of America (H.-H.M. Huang); School of Biological Sciences, Area of Ecology & Biodiversity, Swire Institute of Marine Science, State Key Laboratory of Marine Pollution, University of Hong Kong, Kadoorie Biological Sciences Building, Pokfulam Road, Hong Kong Special Administrative Region (M. Yasuhara).
E-mail addresses: huanghuaihsuan@gmail.com (H.-H.M. Huang), moriakiyasuhara@gmail.com (M. Yasuhara).

https://doi.org/10.1016/j.marmicro.2022.102094
Received 24 January 2021; Received in revised form 4 January 2022; Accepted 19 January 2022
Available online 23 January 2022
0377-8398/

requires the support and collaboration of the scientific community.

Biodiversity databases are rapidly evolving. Traditionally, data have been scattered across platforms in various formats with lower interoperability. One important characteristic of improved biodiversity databases is data mobilization with higher interoperability. The term mobilization refers to making data readily operable for analysis and downloadable through a freely accessible website (Faith et al., 2013; Nelson and Ellis, 2019). It can be applied to all kinds of biodiversity data, including digitized museum collections data, field-based research datasets and laboratory-generated DNA barcode data. The implementation of data mobilization is often consolidated by research infrastructures and digital tools that are designed to (ideally) enhance the reproducibility of scientific studies, to facilitate exchange of datasets in consistently structured ways, to promote open data/science (Burgelman et al., 2019; Powers and Hampton, 2019) and to facilitate integrated data analyses (e.g., Costello and Wieczorek, 2014; Peters and McClennen, 2016; Williams et al., 2018). For example, metadata documentation and many data standards have been developed to allow machine readable exchange of data between databases. Depending on the database, the rules for metadata documentation vary from a free "readme" style to formal standards that define/dictate vocabularies for describing data. Although more work is required, especially for 3D digital morphological data (Davies et al., 2017), communities/working groups such as the Biodiversity Information Standards (TDWG) and the Knowledge Network for Biocomplexity (KNB) have devoted tremendous efforts to maintaining widely used metadata standards in the broad discipline of biology and paleobiology (see below).

A comprehensive introduction to all metadata standards is beyond the scope of this review, but here we briefly mention several widely used examples. The Dublin Core Metadata Initiative defines a set of vocabularies for basic information, such as title and creator (Weibel, 1999). Darwin Core defines a set of vocabularies for biodiversity data based on taxa, such as occurrence and identification (Darwin Core Task Group, 2009; Wieczorek et al., 2012). Audubon Core defines a set of vocabularies for multimedia resources and collections (GBIF/TDWG Multimedia Resources Task Group, 2013; Morris et al., 2013). The Access to Biological Collections Data (ABCD) schema and its extensions establish a data exchange standard for mobilizing data about DNA, specimens, observations and geological samples from museums and botanical gardens (Access to Biological Collections Data Task Group, 2005; Holetschek et al., 2012). Ecological Metadata Language administers a comprehensive set of vocabularies for describing research data and associated details, such as the temporal/spatial extents of data and methods (Fegraus et al., 2005; Jones et al., 2019). These community-maintained metadata standards and their associated software development are the key to mobilizing data from all kinds of data sources.

In this paper, we aim to inform potential database users and contributors of the strengths and weaknesses of all mainstream biodiversity databases that include ostracod records. Ostracods are a group of microscopic crustaceans with an excellent fossil record. They cover a wide variety of ecological niches, have long evolutionary history traced back to the Ordovician (Smith and Horne, 2002; Williams et al., 2008; Yasuhara and Cronin, 2008; Rodriguez-Lazaro and Ruiz-Muñoz, 2012) and are a major component of the meiofaunal biodiversity (Brandt et al., 2007). Being ubiquitous in almost all aquatic systems and commonly preserved as fossils in sedimentary rocks, ostracods are an excellent proxy in contemporary environmental studies, paleoceanographic and paleoclimatic reconstructions, and deep-time paleoecological and macroevolutionary studies (e.g., Holmes and Chivas, 2002; Boomer et al., 2003; Yasuhara and Cronin, 2008; Mesquita-Joanes et al., 2012; Jöst et al., 2019; Yasuhara, 2019; Chiu et al., 2020; Martins et al., 2020). Because of their diverse applications, they have clear potential in the era of open science. Data resources of ostracods have been steadily growing since the pioneering databases, such as the Kempf Database Ostracoda (Viehberg et al., 2014) and Ellis & Messina Catalogues (Ellis and Messina, 1952). Thus, it is timely to evaluate the current role of ostracods in the growing trend of database-oriented research. The aims of this paper are:

- to give an overview of key database projects that involve ostracods, including taxonomically oriented databases (Section 3), Aphia: a multipurpose biodiversity platform (Section 4), marine occurrence-based (paleo)biodiversity databases (Section 5), non-marine occurrence-based (paleo)biodiversity databases (Section 6), and GBIF (Section 8);
- to assess the potential role of ostracod data in understanding the history of life, based on records in the Paleobiology Database (Section 7); and
- to discuss challenges and future directions in developing ostracod databases and their applications (Section 9).

2. Methods

2.1. Database categories

To provide the reader with some orientation in the huge pool of existing databases and tools in general, we have divided them into the following categories (Table 1): (1) taxonomically oriented databases, (2) multipurpose biodiversity databases, (3) data archives/repositories, (4) data harvesters/recombiners, (5) occurrence-based (paleo)biodiversity databases, and (6) application programming interfaces (APIs). The term database can be used to describe a system where datasets are stored, or alternatively to mean a group of datasets. Taxonomically oriented databases focus on authoritative lists of taxonomic names and classification. The multipurpose biodiversity databases category refers to the Aphia platform and its associated databases. Data harvesters/recombiners, such as GBIF (Global Biodiversity Information Facility) and OBIS (Ocean Biodiversity Information System), aggregate data through the Integrated Publishing Toolkit (IPT) and currently host the largest amount of biodiversity data among all the databases discussed herein. The IPT is a free software tool developed by GBIF for registered data publishers, such as the U.S. Geological Survey and the Natural History Museum London, to share biodiversity datasets in the Darwin Core Archive standard (Robertson et al., 2014). Occurrence-based (paleo)biodiversity databases, such as the Arctic Ostracode Database and the Paleobiology Database, rely on authorized scientists or working groups to enter/compile data which mainly come from the published literature. Application Programming Interfaces are not databases, but a general software term to describe an interface system that connects programs and endpoints; they are mentioned here due to their ubiquity and usefulness in data accessibility. Each database has its own unique underlying structure for hosting data. Whereas some database designs may facilitate addressing certain questions better than others, the diversity of designs is complementary and beneficial to almost all lines of research and education (e.g., Wright et al., 2013; Lautenschlager and Rücklin, 2014; Hendricks et al., 2015; Lockwood et al., 2018).

2.2. Reviewed databases

We have reviewed both key ostracod database projects and broader databases that house ostracod data (Fig. 1 and Table 2). For each database, we have studied its stated objectives, data sources, data accessibility, geographical and temporal coverage, and interaction with other databases.

In Section 3, we review the Kempf Database Ostracoda and Ellis & Messina Catalogues that are designed to preserve taxonomic information for both marine and freshwater ostracods. In Section 4, we review the Aphia platform and the World Ostracoda Database (WOD) that house multiple types of data for both marine and freshwater ostracods, including recent and fossil taxa. The Freshwater Animal Diversity Assessment (FADA) project is briefly reviewed in the Aphia platform section because they are in a close collaboration. In Section 5, we review

Table 1
General data resources for (paleo)biology (beyond ostracods). We divide the example databases into six categories according to their characteristics, not for a strict definition but to help the reader navigate through the common data resources. Note that some databases can exchange data with other databases, and thus the same datasets could exist in multiple databases. The web interface of one database can also offer links to multiple databases.

Categories	Examples	Characteristics
Taxonomically oriented databases	ITIS, Open Tree Taxonomy, Catalogue of Life	- Authoritative lists of taxonomic names and classification - Contributed by taxonomic experts - Often as the taxonomic backbone for other databases
Multi-purpose biodiversity databases	Aphia platform (WOD, WoRMS, and ~ 80 registers)	- Authoritative lists of taxonomic names and classification - Contributed by taxonomic experts - Often the taxonomic backbone for other databases - Designed for storing ecological, morphological, biogeographical, stratigraphical and taxonomic data
Data archives /repositories	Dryad, PANGAEA, NOAA's National Centers for Environmental Information, DataOne, Morpho Source (3D datasets), Morphobank (images and phylogenetic matrices), Treebase (phylogenetic matrices)	- Populated by static data in various formats - Contributed by users/authors of publications
Data harvesters /recombiners	GBIF, iDigBio, iDigPaleo, OBIS	- Populated by dynamic data with the formats set by Biodiversity Information Standards (TDWG) - Contributed by registered publishers, such as museums and data repositories, through Integrated Publishing Toolkits
Occurrence-based (paleo) biodiversity databases	Paleobiology Database, Neotoma Database, NOW fossil mammal database, Geobiodiversity Database	- Populated by dynamic data in relational tables - Contributed by verified individual contributors or working groups
Application programming interfaces	Earth Life Consortium, EPANDDA API	- Not a database, but a tool for databases - Linkages between databases or endpoints - A portal for searching data in multiple databases

the Arctic Ostracode Database, Ocean Biogeographic Information System, “Atlas of Atlantic Planktonic Ostracods” and “An Atlas of Southern Ocean Planktonic Ostracods”. An atlas is usually not a database but represents a static output of a database. Nonetheless, we include these two atlases here because they are valuable resources for planktonic ostracods, and because the original data have been partially published by the Natural History Museum London in OBIS as a dataset called “Personal library collection of Martin Angel of published and unpublished Halocyprid (Ostracoda) occurrences” (Angel, 2016). In Section 6, we review non-marine occurrence-based (paleo)biodiversity databases. In Section 7, we visit ostracods in the Paleobiology Database (PBDB), one of the most comprehensive spatiotemporal paleoecological databases that accommodate fossil data (Ordovician to Quaternary for ostracods). In Section 8, we introduce GBIF, which aggregates data from all above databases through Integrated Publishing Toolkits.

We do not list static individual datasets deposited in data archives/repositories, such as PANGAEA and NOAA's National Centers for Environmental Information. This is because many mainstream data repositories have joined data harvesters, such as OBIS and GBIF. In Sections 5.2. and 8, we consider how OBIS and GBIF use metadata standards and recombine datasets from numerous data repositories, respectively. Among these data repositories, PANGAEA and ANTABIF are introduced to illustrate the ostracod data in OBIS.

Neptune Sandbox Berlin curates microfossil occurrence and biostratigraphy data from the DSDP/ODP/IODP projects, but include only fossil microplankton (e.g., calcareous nannofossils, foraminifera, radiolarians, diatoms, and dinoflagellates) (Lazarus, 1994). Thus, it is excluded from our review since planktonic ostracods (and other planktonic arthropods) rarely become microfossils due to their poorly calcified carapace (Perrier et al., 2015). However, it should be mentioned that numerous studies have generated excellent fossil records of benthic ostracods from DSDP/ODP/IODP samples (e.g., Majoran and Dingle, 2001; Bergue and Govindan, 2010; Alvarez Zarikian, 2015).

Abbreviations: Antarctic Biodiversity Information Facility (ANTABIF), Arctic Ostracode Database (AOD), Australian Non-marine Ostracode Database (AUNODe), Freshwater Biodiversity Data Portal (BioFresh), Catalogue of Life (CoL), Chinese Non-marine Ostracode Database (CHINODe), Delorme Ostracode Autecological Database (DOAD), East Asia Non-marine Ostracod Database (EANODe), Encyclopedia of Life (EoL), EU-FP6 project Marine Biodiversity and Ecosystem Functioning Network of Excellence (MarBEF), Ellis & Messina Catalogues (E&M), Freshwater Animal Diversity Assessment (FADA), Global Biodiversity Information Facility (GBIF), Kempf Database Ostracoda (KDO), Nonmarine Ostracod Distribution in Europe (NODE), North American Combined Ostracode Database (NACODe), North American Nonmarine Ostracode Database (NANODe), Ocean Biogeographic Information System (OBIS), Ostracod Metadatabase of Environmental and Geographical Attributes (OMEGA), Paleobiology Database (PBDB), South African Non-marine Ostracode Database (SANODe), World Register of Marine Species (WoRMS), World Ostracoda Database (WOD).

2.3. Data analysis

We calculated database status metrics where possible, including spatial coverage, temporal coverage, resolution of taxonomic identification, and database growth over time. Table 3 summarizes the download information, the requested taxa, and parameters applied on the raw data for each analysis. In Section 5.1., although a detailed biogeographical analysis is beyond the scope of this paper, we applied a network-based clustering analysis on the census data from the Arctic Ostracode Database to show a simple example of what type of analyses this database allows. This network-based clustering analysis determines biogeographical clusters or regions by calculating number of co-occurrences between pairs of species and number of shared species between pairs of sites in a bipartite occurrence network (Vilhena and Antonelli, 2015; Kocsis et al., 2018a). We used only samples with >100 specimens in the network analysis. In Section 7, we calculated the raw number of occurrences and genera in Ostracoda per geological period in the Paleobiology Database. We also calculated the number of genera per order in Ostracoda, including Podocopida, Palaeocopida, Platycopida, and Myodocopida.

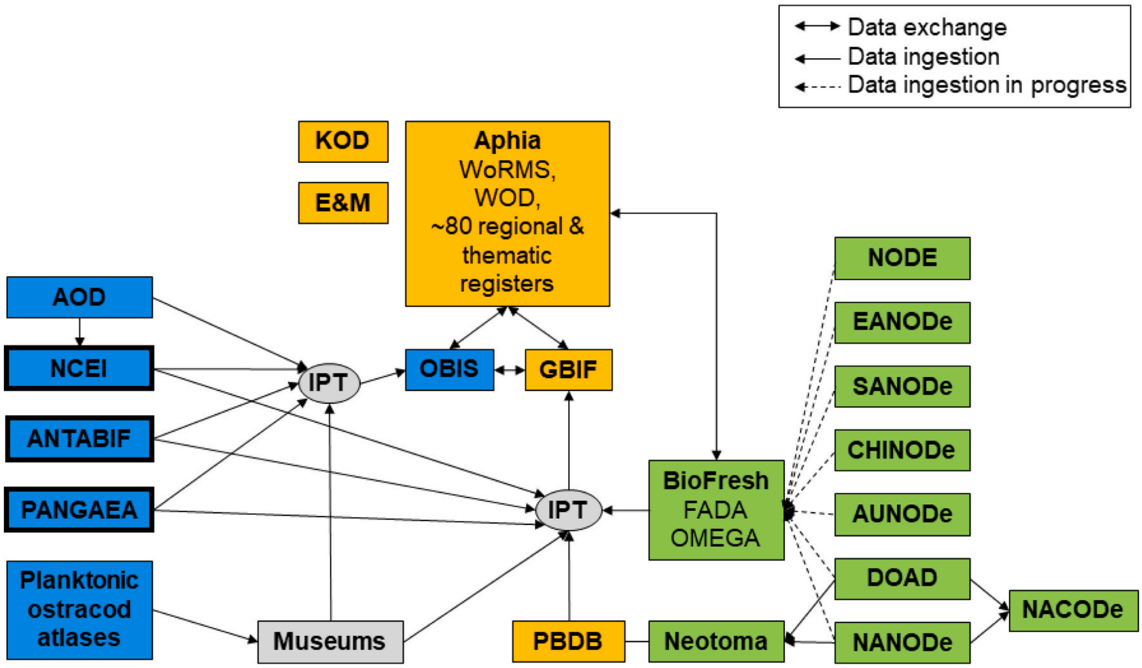

Fig. 1. Diagram of the relationship among reviewed databases (in the broadest sense). This diagram shows that many databases are in partnership rather than in isolation. Marine databases (blue) and non-marine databases (green) are connected through databases (orange) that accommodate both information. Some databases (bold) are independent, and some (not bold) are a part of or adopted into Aphia or BioFresh. We simplified the connections to three kinds: (1) data exchange (solid double arrows) between independent databases, (2) data ingestion from regional databases to broader-scale databases (solid arrows), and (3) data ingestion in progress (dashed arrows). Paleobiology Database (PBDB) and Neotoma Database are in close collaboration (a solid line) through Earth Life Consortium. GBIF and OBIS aggregate data through Integrated Publishing Toolkits (gray circles); NCEI, ANTABIF, and PANGAEA (thickened box border) are contributed by authors of publications; other databases are maintained by authorized editors/enterers/data stewards/working groups. Specimens and collections from museums (gray) are important data sources to GBIF and OBIS. "Planktonic ostracod atlases" refer to "Atlas of Atlantic Planktonic Ostracods" and "An Atlas of Southern Ocean Planktonic Ostracods". Abbreviations: see Section 2.2. for full names. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Taxonomically oriented databases

The Kempf Database Ostracoda (KDO) is one of the earliest initiatives to compile taxonomic indexes and references. Kempf objectively compiled citations for more than 40,000 marine and 9000 non-marine taxonomic names of living and fossil species from the Ordovician to the present day (Matzke-Karasz, 2014). It is the single most complete and comprehensive listing of ostracod taxonomic names available, referenced to the publications in which taxa were originally described, and constituting a unique and valuable resource. It was published in CD and book forms by the University of Cologne; the data do not include images, descriptions, or synonymies. Since Kempf's death in 2017 there has been a necessary pause in the entry of new data while efforts are made to establish support for the KDO and thus ensure its future. It is planned to make it widely available, for example via the German node of the Global Biodiversity Information Facility (GBIF), provided that funds can be obtained to facilitate this.

Ellis & Messina Catalogues (Ellis and Messina, 1952; Micropaleontology Press, 2021) reproduce type descriptions and illustrations of genus- and species-level taxa, together with details of stratigraphical and geographical locations and the repositories of type material. New information beyond the original description, such as the repository of type material that is not shown in the original description and taxonomic note, is available for some species and genera. In addition to ostracods (>26,000 taxa) the catalogues cover foraminifera (>47,000 taxa) and diatoms (>7,000 taxa), with around 300 new taxa being added annually to each catalogue (Micropaleontology Press, 2021).

The World Ostracoda Database (WOD) was established as a taxonomically oriented database in its first phase of database construction, compiling taxonomic information on fossil and Recent/living marine and non-marine ostracods. Using the robust infrastructure of the Aphia platform, the taxonomic information is accessible online and constantly growing. Ongoing further development phases include adding a wider range of data (e.g., biogeographical, ecological, evolutionary) (see Section 4. for more details).

4. Aphia, a multipurpose biodiversity platform, and the World Ostracoda Database

The European Register of Marine Species (ERMS) is an authoritative (i.e., made by taxonomic experts) list (initially published as a printed book) of >29,000 (geologically) Recent marine and brackish-water species from European waters, of which some 700 are ostracods, together with information on their taxonomy and references for their taxonomic identification (Costello et al., 2001; Cuvelier et al., 2006). ERMS can now be accessed via the MarBEF web site (MarBEF, 2004). Through a series of mostly EU funded programs, projects, initiatives, and institutions (e.g., MarBEF, Species 2000 Europa, Flanders Marine Institute, Species 2000), ERMS was fed into a relational digital database (Cuvelier et al., 2006), which later became the Aphia platform. ERMS was then expanded to a global register, the World Register of Marine Species (WoRMS), and later the initiative expanded further to compiling information on all living and fossil species from all ecosystems (marine and non-marine), as well as starting to incorporate all aspects of taxonomy and nomenclature, also biological, ecological, evolutionary, biogeographical, genetic, bibliographical and nomenclatural information, conservation importance, economic importance, images, and notes (e.g., stratigraphy, taxonomic remarks, description). All these data are displayed to the public in the form of taxonomic, regional, and thematic registers, each one with its own webpage. Examples of these three kinds

Table 2
Data resources reviewed in this paper. The table is arranged by the appearance sequence of the databases in Sections 3–8. The network column lists some databases that link with the focal database.

Name	Category	Marine	Non-marine	Age accommodation	Network	Reference
Kempf Database Ostracoda (KDO)	Taxonomically oriented database	√	√	Extant and extinct	–	(Viehberg et al., 2014)
Ellis and Messina Catalogues: Ostracoda	Taxonomically oriented database	√	√	Extant and extinct	–	(Ellis and Messina, 1952; Micropaleontology Press, 2021)
(Aphia) World Ostracoda Database (WOD)	Multipurpose biodiversity database	√	√	Extant and extinct	EoL, ITIS, CoL, GBIF, OBIS, etc.	(Brandão et al., 2022)
Freshwater Animal Diversity Assessment (FADA): Ostracoda	Taxonomic checklists generated from an occurrence-based (paleo) biodiversity database		√	Extant	Freshwater Information Platform, BioFresh, Aphia, GBIF	(Martens et al., 2008; Martens et al., 2013)
Arctic Ostracode Database (AOD)	Occurrence-based (paleo) biodiversity database	√		Modern to the Quaternary	NOAA's NCEI, OBIS	(Cronin et al., 1995; Gemery et al., 2017; Cronin et al., 2021)
Ocean Biogeographic Information System (OBIS)	Occurrence-based data harvester/recombiner	√		Modern to the Quaternary	GBIF, PANGAEA, ANTABIF	(OBIS, 2021)
PANGAEA	Data archive/repository	√		Modern to the Quaternary	OBIS, GBIF	(PANGAEA®, 2021)
Antarctic Biodiversity Information Facility (ANTABIF)	Data archive/repository	√		Modern/Recent	Antarctic Biodiversity Portal, OBIS	(biodiversity.aq, 2021)
Atlas of Atlantic Planktonic Ostracods	Outputs generated from an occurrence-based (paleo) biodiversity database	√		Modern/Recent	Natural History Museum London, OBIS	(Angel et al., 2008)
An Atlas of Southern Ocean planktonic ostracods	Outputs generated from an occurrence-based (paleo) biodiversity database	√		Modern/Recent	OBIS	(Blachowiak-Samolyk and Angel, 2008)
Non-marine Ostracod Distribution in Europe (NODE)	Occurrence-based (paleo) biodiversity database		√	Modern to the Quaternary	OMEGA	(Horne et al., 2011)
North American Non-marine Ostracode Database (NANODe)	Occurrence-based (paleo) biodiversity database		√	Modern/Recent	Neotoma Database, OMEGA, NACODe	(Smith et al., 2015)
Delorme Ostracode Autecological Database (DOAD)	Occurrence-based (paleo) biodiversity database		√	Modern/Recent	Neotoma Database, OMEGA, NACODe	(Curry et al., 2012)
East Asia Non-marine Ostracod Database (EANODe)	Occurrence-based (paleo) biodiversity database		√	Modern/Recent	OMEGA	Unpublished compilation
Ostracod Metadatabase of Environmental and Geographical Attributes (OMEGA)	Occurrence-based (paleo) biodiversity database		√	Modern/Recent	Freshwater Information Platform, BioFresh, GBIF	(Horne et al., 2011)
Neotoma Database	Occurrence-based (paleo) biodiversity database		√	Modern to Pliocene	Earth Life Consortium, NCEI	(Williams et al., 2018)
North American Combined Ostracode Database (NACODe)	Occurrence-based (paleo) biodiversity database		√	Modern/Recent	–	(Curry et al., 2012)
Paleobiology Database (PBDB)	Occurrence-based (paleo) biodiversity database	√	√	Phanerozoic	Earth Life Consortium, ePANDDA API, GBIF	(Peters and McClennen, 2016)
Global Biodiversity Information Facility (GBIF)	Data harvester/recombiner	√	√	Living and fossil	OBIS, Paleobiology Database, Biofresh, etc.	(GBIF, 2021)

of registers are, respectively, the World Ostracoda Database (WOD), the Register of Antarctic Species (RAS) and the World Register of Introduced Marine Species (WRiMS). The different registers' web interfaces make the information contained in Aphia available to the public in a more accessible way. For example, someone looking for data on ostracods will tend to access the WOD webpage, while someone looking for Antarctic species will access RAS. However, all data are housed within Aphia, and the web interfaces provide search tools which communicate with the Aphia database.

Since its inception Aphia grew significantly and several tools were created; one worthy of mention is the online editing environment available since 2006, through which hundreds of accredited editors can log into the system and edit or add information only to their specific taxa (Costello, 2000; Vandepitte et al., 2015). In this way, although the information in the database is accessible to the general public, the quality of the data is assured by the fact that only accredited specialists can edit the information in the database. For example, only an editor can add, delete and correct taxonomic, ecological, evolutionary, biogeographical, genetic, bibliographical and nomenclatural information, conservation importance, economic importance, images, stratigraphical and other kinds of notes, etc. (Vandepitte et al., 2015). Furthermore, mistakes and omissions can be constantly detected and passed to and corrected by the responsible editors. However, the lack of editors responsible for specific subjects or taxa will surely delay the correction of such mistakes. The low number of editors is now the main problem of WOD, with the consequence that mistakes take longer to be corrected. Nevertheless, a clear advantage of these open access, online databases (such as WOD) is to simplify or avoid taxonomic confusion, e.g., synonyms and homonyms. One example in ostracodology is *Vandenboldina* Wilson, 2010 published as a new name for *Pseudoceratina* Bold, 1965, after Simone N. Brandão (editor-in-chief of the WOD) advertised it via Ostracon (the email list for ostracod researchers) as a junior homonym of *Pseudoceratina* Carter, 1885 (Porifera). Although the first objective of ERMS was to produce a species list, a much broader idea lay behind its initial proposals (e.g., Costello, 2000, Fig. 3), encompassing networking with other databases and scientific programs (for example, oceanographic

Table 3
Data sources used in this paper. Data source includes the download information, request parameters, web-link, and access date.

Figures	Data sources
Fig. 2	All data downloaded from the publicly accessible Arctic Ostracode Database 2020 (Cronin et al., 2021) in National Centers for Environmental Information (https://www.ncei.noaa.gov/) on 2021-10-13.
Fig. 3	All ostracod data downloaded from the publicly accessible Ocean Biogeographic Information System (https://obis.org/) on 2020-10-17.
Fig. 4	All data provided by David J. Horne in December 2020, from Australian Non-marine Ostracode Database (AUNODe), Chinese Non-marine Ostracode Database (CHINODe), Delorme Ostracode Autecological Database (DOAD), East Asia Non-marine Ostracod Database (EANODe), North American Non-marine Ostracode Database (NANODe), Non-marine Ostracod Distribution in Europe (NODE) database, South African Non-marine Ostracode Database (SANODe)
Fig. 5–7	All data downloaded from the publicly accessible Paleobiology Database (https://paleobiodb.org/) on 2020-10-11. Requested taxa: Arthropoda. Applied filters: taxonomic resolution (all), preservation (regular taxa only), identification (latest), show accepted names only.

atlases with information on water masses, bathymetry, nutrients, etc.), online tools for capturing, storing and displaying biogeographical information, and modelling of future species distribution and ocean ecosystem functions.

The Aphia platform contains a database, which hosts all data from WOD, WoRMS and the remaining ~80 global, regional, and thematic registers. The SQL database of Aphia with "over 400 fields, spread over 81 related tables", which are grouped in relation to the content in "10 modules: taxonomy, distribution, traits, specimen information, vernacular names, notes, links, images, identification keys and sources" (Vandepitte et al., 2015). The EU-funded Lifewatch infrastructure for biodiversity and ecosystem research is promoting the data mining and upload of a large amount of biological, morphological, evolutionary, ecological, taxonomic and human defined characteristics of virtually all taxa, including ostracods, to the module "traits" of Aphia, e.g., body size, feeding type, life cycle, invasive species, threatened species, etc. (Vandepitte et al., 2015). Aphia provides data to other databases (e.g., GBID, OBIS, CoL, ITIS, several museums, scientific institutions and individual scientists) and provides links to them in the specific taxa pages, in order to, for example, avoid duplication of work (Cuvelier et al., 2006; Vandepitte et al., 2015).

Within the scope of the AquaRES (Aquatic Species Register Exchange and Services) project, Aphia exchanges data with the Freshwater Animal Diversity Assessment (FADA). The AquaRES project aims at managing the data of aquatic groups occurring in both freshwater and marine ecosystems in one system (Vandepitte et al., 2015). The FADA database, led by Koen Martens from the Royal Belgian Institute of Natural Sciences, provides taxonomic checklists and distribution information of many freshwater animals, including ostracods (Balian et al., 2008; Martens et al., 2013). Georeferenced datasets can be made available via the Freshwater Information Platform which integrates FADA through the Freshwater Biodiversity Data Portal. In an output of FADA, non-marine ostracods are grouped into eight biogeographic regions: Antarctic (3 species), Australasian (262 species), Afrotropical (453 species), Nearctic (300 species), Neotropical (290 species), Oriental/Indomalaya (222 species), Pacific/Oceania (57 species) and Palaearctic (749 species) (Martens et al., 2013). The distribution of each taxon is thus limited to its presence anywhere within a named biogeographical region, rather than single occurrences identified by coordinates and other data (e.g., latitude, longitude, altitude, date of sampling). The taxonomic classification of each taxon is provided, but not the taxonomic history or original description.

The World Ostracoda Database (WOD) is one of the global species registers housed in Aphia. Its aims are to compile a wide range of data types on recent and fossil Ostracoda, including taxonomy, classification, synonymy, description, bibliography, type specimens' information, geographical and stratigraphical occurrences, illustrations, biological and ecological attributes (e.g., feeding type, body size, functional group, development, life cycle, habitat). WOD originated in 2008, when the chair of the WoRMS, Geoff Boxshall invited Simone N. Brandão to check a spreadsheet with approximately 8,000 ostracod taxa compiled from publications on Recent marine Ostracoda. In 2013, a grant from LifeWatch enabled a large input of data on Ostracoda, including tens of thousands of taxa and bibliographic references on marine, non-marine, extant, and extinct ostracods. WOD has kept expanding since then and has information now on more than 54,449 taxa: 45,294 species, of which 32,140 are accepted, and 3,676 genera, of which 3,446 are accepted.

WOD, WoRMS and the remaining registers in Aphia aim to record the entire taxonomic history of a taxon, and thus unaccepted taxa are kept in the database with an "unaccepted" label. In most cases these unaccepted taxa are invalid generic combinations at the species level, but other unaccepted taxa are subjective synonyms (i.e., taxa described with a different name and authorship, but considered later to be a synonym of a taxon described before). However, most taxa now available in WOD remain quarantined (hidden from the public). These taxa could be the ones without higher level taxonomic classification or those not yet checked by an expert. Nevertheless, 10,221 accepted species among a total of 20,548 taxa are available to everyone on the worldwide web (Brandão et al., 2022), the remaining +35,000 species in WOD will be checked before they are made accessible to the public. WOD also contains the citations of +22,000 publications on ostracods and + 2,000 pdfs, which can be downloaded either directly (if no copyright applies) or upon request to the WOD (if the publication is copyrighted). Furthermore, WOD has also +2,200 distribution records, +700 images and ~ 800 attributes (e.g., biological, ecological, developmental, evolutionary data). In the context of Lifewatch Project, ongoing steps include entering biogeographical, ecological, evolutionary data for different taxa, as well as keeping the taxonomic information and the references up to date.

5. Marine occurrence-based (paleo)biodiversity databases

5.1. Arctic Ostracode Database

The Arctic Ostracode Database (AOD) is dedicated to census data of benthic marine ostracods collected from a variety of surface and late Quaternary sediment samples from the Arctic area (Fig. 2). The samples came from international sampling cruises to the Arctic since 1933 (Fig. 2B). The census data provide numbers of specimens per species in each sample. It also provides taxonomic information of Arctic ostracods, which all the contributors agreed on. Therefore, taxonomic names are thoroughly harmonized in the AOD. Age information at finer resolution than "modern to late Quaternary" is unavailable in the AOD but might be indicated by papers based on the same samples elsewhere.

The development of the AOD has been led by Thomas M. Cronin and Laura Gemery in the United States Geological Survey (USGS). The home repository of the AOD is NOAA's National Centers for Environmental Information. It started as the Modern Arctic Podocopid Ostracode Database (Cronin et al., 1991) and had several major updates, including Arctic Ostracode Database (Cronin et al., 1995), Modern Arctic Ostracode Database (Cronin et al., 2010), Arctic Ostracode Database-2015 (Gemery et al., 2017), and Arctic Ostracode Database-2020 (Cronin et al., 2021) (Fig. 2B). The online USGS version of the Arctic Ostracode Database was contributed by Thomas M. Cronin, Thomas R. Holtz, Elisabeth M. Brouwers, William M. Briggs, Robin C. Whatley and Adrian Wood (Cronin et al., 1995). Some of the datasets in the AOD are also present in other databases, such as PANGAEA and OBIS. Through the Integrated Publishing Toolkit, OBIS (OBIS Secretariat, 2021) also harvested the Modern Arctic Ostracode Database (Cronin et al., 2010) and the Arctic ostracod datasets in PANGAEA database (see Section 5.2.). Note that the census counts are flattened to occurrence data (i.e.,

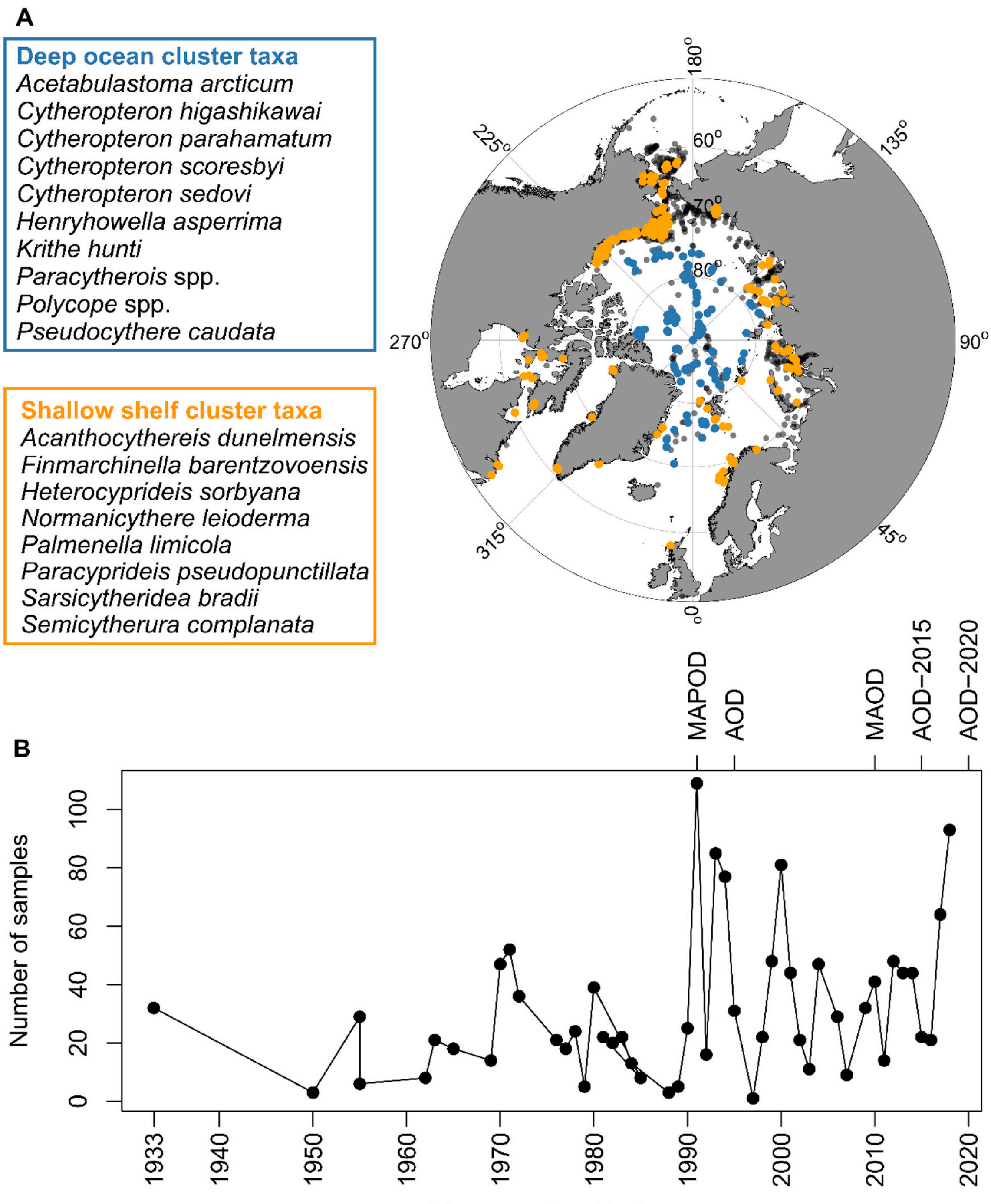

Fig. 2. Arctic Ostracode Database (AOD): Major Arctic taxa and the sampling history. (A) Two clusters determined by the network-based clustering analysis. The polar deep ocean taxa (listed in the blue box) are relatively more abundant and frequently occur in the deep ocean cluster (blue points on the map), and the same applies to the shallow shelf taxa (in the orange box) in the shallow shelf cluster (orange points on the map). Note that *Acetabulastoma arcticum* is a parasitic species living on sea-ice dwelling amphipods (Cronin et al., 2010). Only samples with more than 100 specimens in the AOD-2015 version were used in the clustering analysis; the black points are the AOD samples that are not used in the analysis. (B) Number of samples sorted by the year of sampling. In parallel to the x-axis, the AOD versions are indicated by their published year: Modern Arctic podocopid ostracode database (MAPOD in Cronin et al., 1991); Arctic Ostracode Database (AOD in Cronin et al., 1995); Modern Arctic Ostracode Database (MAOD in Cronin et al., 2010); Arctic Ostracode Database-2015 (AOD-2015 published online in 2015 (Gemery et al., 2017)); Arctic Ostracode Database-2020 (Cronin et al., 2021). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

presence/absence) in the OBIS.

The latest version, AOD-2020, incorporates the previous versions and provides georeferenced occurrence and species census data for 96 species of benthic marine ostracods from over 1500 modern surface sediment samples (Cronin et al., 2021; Gemery et al., 2021a; Gemery et al., 2021b). This unique database grants us an opportunity to analyze and understand Arctic ostracods beyond the scale of a local project. To gain an overview, we applied a network-based clustering analysis to the AOD-2015 census data (see Gemery et al., 2017 for detailed species information and biogeography). The results suggest two biogeographical clusters (Fig. 2A): a deep ocean cluster is distributed in the open ocean at latitudes of 70°N and higher, and a shallow shelf cluster in the shallow marine coastal areas at latitudes of 80°N and lower.

5.2. Ocean Biogeographic Information System

The Ocean Biogeographic Information System (OBIS) is dedicated to occurrence records, event records, and measurements/facts of modern marine life. It started as one initiative of the Census of Marine Life (2000–2010) and had been adopted into the UNESCO's International Oceanographic Data and Information (IODE) program since 2010. OBIS gathered data through Integrated Publishing Toolkits from more than 20 OBIS regional/country/thematic nodes (e.g., Arctic OBIS, OBIS Germany, Fish OBIS), which connect hundreds of institutional data publishers or data archives (e.g., PANGAEA, NOAA's NCEI, USGS). The Integrated Publishing Toolkit, developed by GBIF, assists with formatting the data to valid Darwin Core terms and describing dataset metadata in Ecological Metadata Language. This readily allows interoperability of datasets. OBIS reached 0.5 million records in 2002, and over 63 million occurrence records in 2020 (obis.org), together with geographic and bathymetric distribution and environmental data (e.g., salinity, temperature).

To ensure the utility of a system such as OBIS, which integrates datasets from multiple sources that were originally compiled in a variety of circumstances and for different purposes, a certain amount of data and metadata standardization is essential (OBIS, 2021). This may be achieved by having mandatory fields that must be completed for the dataset to be accepted. For example, biogeographical data in OBIS requires locations to be defined by latitude and longitude coordinates in decimal degrees to facilitate mapping of records, so inclusion of datasets that use a different coordinates system such as Universal Transverse Mercator (UTM) requires conversions. Another example is taxonomic harmonization, which in the case of OBIS is achieved through the matching of names to authoritative taxonomic lists such as the WoRMS (see Section 4).

OBIS focuses on modern marine life, but its ostracod data include both modern records from various types of samples and Quaternary records from core section sediments. At the time of writing in 2020, it houses nearly 205,000 records for 2,787 ostracod taxa (2,415 species), including both benthic and planktonic taxa. WoRMS provides the taxonomic backbone to OBIS, and about 49.4% of ostracod records were identified to species level in OBIS (Fig. 3C). Ostracod data show global coverage in OBIS, but there is a strong disparity in sampling efforts between different oceans and between coastal and open oceanic regions. The areas of best coverage concentrate mostly in: (1) the northern hemisphere, and particularly the North Atlantic (versus the southern hemisphere and the Pacific Ocean; Fig. 3A); (2) coastal regions (versus open ocean and deep ocean; Fig. 3B).

PANGAEA is an important contributor to the ostracod data in OBIS. Hosted by the Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI) and the Center for Marine Environmental Sciences (MARUM), PANGAEA is a mainstream data archive/repository for georeferenced data from earth system research, e.g., Deep Sea Drilling Program (DSDP, 1967–1983), Ocean Drilling Program (ODP, 1984–2002), Integrated Ocean Drilling Program (IODP, 2003–2013), the International Ocean Discovery Program (IODP, 2013–2023), and many other research projects of various scales. Despite its focus on modern marine ecosystems, OBIS also harvested paleontological datasets from PANGAEA.

Endorsed by the Scientific Committee on Antarctic Research (SCAR), the Antarctic Biodiversity Information Facility (ANTABIF) exchanges biogeographical data of marine species with the Antarctic Thematic Node of OBIS, the AntOBIS, and provides data ultimately to GBIF. ANTABIF is a complex platform focused on continental and marine research in/off Antarctica, and it houses a diverse collection of

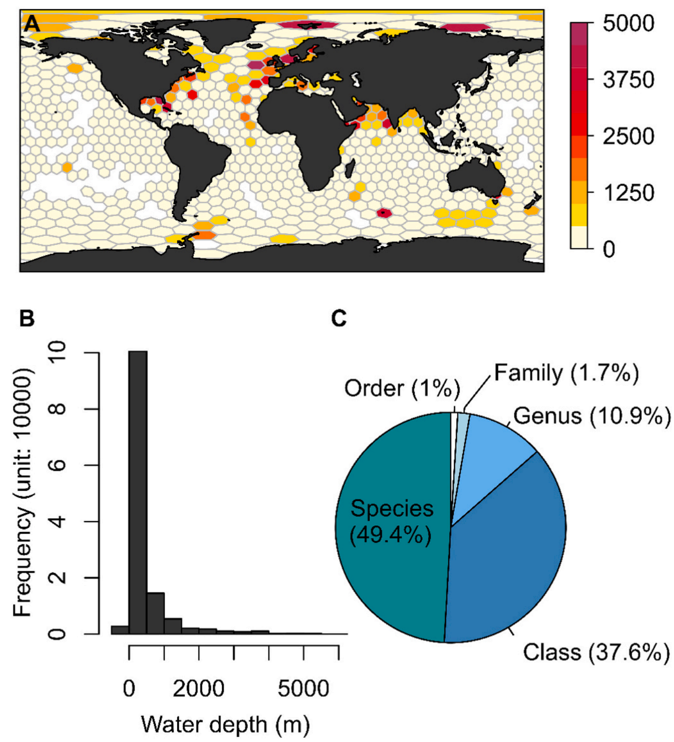

Fig. 3. Ocean Biogeographic Information System (OBIS): global records of marine and brackish ostracods. (A) Geographical distribution of the records in a heat-map style. (B) Water depth distribution of the records. (C) Identification resolution of the records. The map shows the discrepancies in number of records across oceans in the OBIS. Note the intensely studied regions: European and North American waters; and the generally understudied areas: Pacific, Indian, and Southern Oceans. The Arctic Ocean is relatively rich in ostracod data which were harvested from PANGAEA and the Arctic Ostracode Database (see Section 5.1.). Most records come from 0 to 500 m water depths. Most records (49.4%) had identification resolution down to the species level, but many records (37.6%) were just "Ostracoda".

oceanographical, ecological, biological and geological datasets (e.g., SCAR Southern Ocean Diet and Energetics Database), as well as providing online analysis tools for the scientific community (e.g., interactive identification keys, R packages relevant to Antarctic and Southern Ocean science). Concerning the biogeographical data exchanged with OBIS, currently ANTABIF hosts over 190 datasets of many taxonomic groups, including Southern Ocean Ostracoda (excluding Halocypridina) (Brandão, 2012), which focuses on benthic ostracods. These datasets can be accessed through the Antarctic Biodiversity Portal (biodiversity.aq, 2021), which integrates datasets compiled under the SCAR-MarBIN (SCAR-Marine Biodiversity Information Network) initiative with data from different sources, including the Australian Antarctic Division and other institutions. The Southern Ocean benthic Ostracoda dataset (Brandão, 2012) includes 888 occurrences of 113 taxa from 193 georeferenced locations. Planktonic Southern Ocean ostracods were covered by one atlas, which offer the taxonomy, geographical distribution (static occurrence maps), illustrations and measurements of key morphological features of species (mostly Halocypridina, but all known planktonic Cypridinida are also present; see the next section).

5.3. Marine planktonic ostracods

Martin Angel, Kasia Blachowiak-Samolyk, and Vladimir Chavtur compiled the available information on marine planktonic ostracods from the Atlantic and Southern Oceans, and published the "Atlas of Atlantic Planktonic Ostracods" (Angel et al., 2008) and "An Atlas of Southern

Ocean Planktonic Ostracods" (Blachowiak-Samolyk and Angel, 2008). The Atlantic compilation (Angel et al., 2008) includes data from before the Challenger expedition (1870s) up to 2007, while the Southern Ocean dataset (Blachowiak-Samolyk and Angel, 2008) include data from the Discovery Investigations (1930s) up to 2007. These two compilations include detailed taxonomic and morphological (e.g., size, shape) data, static maps of geographical distribution, illustrations and measurements of important morphological characters, and list of bibliographies, all in pdf format. This means that although these are valuable resources on planktonic ostracods, specific data on the occurrences (latitude, longitude, depth, date of collection, etc.) cannot be extracted electronically from these atlases, neither it is possible to exchange data in a machine-readable format.

6. Non-marine occurrence-based (paleo)biodiversity databases

There are several regional databases focused on the ecology and distribution of living non-marine ostracods. Prominent among these are NODE (Non-marine Ostracod Distribution in Europe), NANODe (North American Non-marine Ostracode Database), DOAD (Delorme Ostracode Autecological Database), and EANODe (East Asia Non-marine Ostracod Database); key metadata and data from these and other regional databases are compiled in OMEGA (Ostracod Metadatabase of Environmental and Geographical Attributes). Datasets from DOAD and NANODe have been amalgamated into a North American Combined Ostracode Database (NACODe) as well as being made accessible through the Neotoma Paleoecology Database. Details of each of these are described briefly below. Distribution maps of all records from the above databases are in Fig. 4.

6.1. Non-marine Ostracod Distribution in Europe

The Non-marine Ostracod Distribution in Europe (NODE) database contains approximately 10,000 records of living ostracod species (plus about 2,000 Pleistocene and Holocene fossil records), representing more than 400 species and approximately 2,500 localities. It was initiated by six research teams in five countries to map the distribution of ostracod sex and parthenogenesis for the three-year (1994–1996) EU Human Capital and Mobility Programme project *Evolutionary ecology of reproductive modes in non-marine Ostracoda* (Horne et al., 1998) and subsequently developed further by David J. Horne as its Data Steward (i.e., the person responsible for all aspects of database management including quality control and accessibility). Focused on Europe, its geographical coverage extends longitudinally from the Azores to east of the Black Sea and latitudinally from the Canary Islands to Svalbard (approx. 32°W – 45°E and 25–80°N). A few additional North American and Asian records are currently included in NODE for calibration purposes in connection with Quaternary paleoclimatic reconstruction (see discussion below); these may eventually be transferred to other databases. The records in NODE, compiled almost entirely from scientific literature published in the 19th, 20th and 21st centuries, are focused on occurrences of identified ostracod species but include (where possible) information such as water body type, date of collection, and whether males were found.

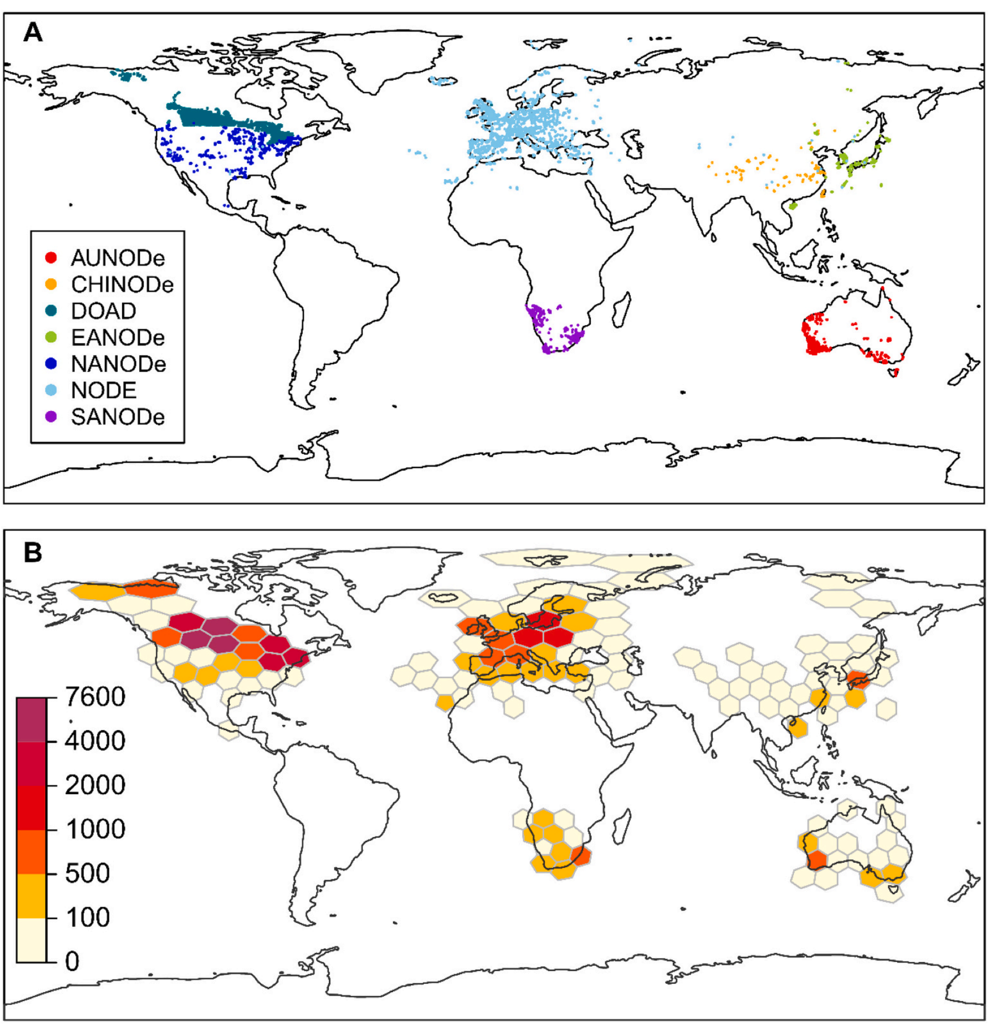

Fig. 4. Non-marine occurrence-based (paleo)biodiversity databases. (A) Geographical distribution of all the sites in the Ostracod Metadatabase of Environmental and Geographical Attributes (OMEGA), colour-coded according to the contributing databases: Australian Non-marine Ostracode Database (AUNODe), Chinese Non-marine Ostracode Database (CHINODe), Delorme Ostracode Autecological Database (DOAD), East Asia Non-marine Ostracod Database (EANODe), North American Non-marine Ostracode Database (NANODe), Non-marine Ostracod Distribution in Europe (NODE) database, South African Non-marine Ostracod Database (SANODe). (B) A heatmap of all occurrence records from (A); note the higher concentrations of records in North America and Europe.

6.2. North American Non-marine Ostracode Database

The North American Non-marine Ostracode Database (NANODe) was initiated by Richard M. Forester at the U.S. Geological Survey. Field collection efforts were conducted primarily by Forester et al. (2005), and data subsequently managed by Alison J. Smith as Data Steward. The dataset (www.kent.edu/nanode) is composed of late 20th to early 21st century collections (1979 through 2005) and contains approximately 2,600 records comprising *c.* 100 species and *c.* 600 localities with accompanying major ion hydrochemical and field limnological data collected at the same time as the species collection (Forester et al., 2005). It provides a well-distributed coverage of the conterminous states of the USA from California in the west to New Hampshire in the east (approx. 124–70^{o}W and 29–49^{o}N) although the southeastern states (Florida to the Carolinas) are not yet represented. The distribution of each of the species in NANODe is shown in geographic maps and in solute graphs (www.kent.edu/nanode). The data are completely accessible through the multiproxy Neotoma Paleoecology Database (see below). The slide collection is housed in the Geology Department at Kent State University, Kent, Ohio (USA).

6.3. Delorme Ostracode Autecological Database

The Delorme Ostracode Autecological Database (DOAD), based on primary collection by Denis Delorme during the 1960s and 1970s, includes over 30,000 records representing *c.* 130 species and more than 6,000 localities, together with extensive geographical, environmental, and climatic data for the waterbodies sampled. Both the database and a very substantial collection of voucher specimens (representing almost every record in the database) are now housed and curated by the Canadian Museum of Nature at its Natural Heritage Campus (Research and Collections Facility) in Gatineau, Canada, together with a collection of type and illustrated ostracod specimens (related to Delorme's publications) that was formerly kept at the Geological Survey of Canada in Ottawa. A high density of records covers Canada south of latitude 60^{o}N from southeastern British Columbia in the west to the Great Lakes region and as far as southwestern Quebec in the East (approx. 120–70^{o}W and 41–60^{o}N); there are also records in Yukon Territory and Northwest Territories (approx. 140 –128^{o}W and 65–70^{o}N).

6.4. East Asia Non-marine Ostracod Database

The East Asia Non-marine Ostracod Database (EANODe), compiled and managed by Robin J. Smith as its Data Steward, contains approximately 1,700 records representing c. 150 species and c. 650 localities, of which the majority are in Japan but including South Korea, China and the Far Eastern Federal District of Russia (approx. 108–145°E and 18–72^{o}N). The data are mainly from primary collections represented by specimens curated at the Lake Biwa Museum in Japan (c. 75%), the remainder being compiled from published literature (25%).

6.5. Ostracod Metadatabase of Environmental and Geographical Attributes

A single global database of non-marine ostracod distribution would be of immense scientific value, but creating one would be a very challenging, long-term project. As a more pragmatic approach, with short-term as well as long-term benefits, we initiated the OMEGA: Ostracod Metadatabase of Environmental and Geographical Attributes (Horne et al., 2011), with the aim of compiling and maintaining a metadatabase of regional non-marine ostracod databases.

It is important to understand that OMEGA is not a merging of databases to form a "super-database". A metadatabase contains data about databases, and in this sense OMEGA might simply be a database of existing databases that contain information on ostracod ecology and distribution in different parts of the world. However, the most important data (taxonomic names and coordinates of records) are included as metadata, thus facilitating the mapping and searching of records from all of the contributing databases. Each record in OMEGA is attributed to a source database for which contact details will be supplied. It is hoped that, ultimately, links could be included to facilitate access to the regional databases; a pre-requisite for such a functionality, however, is that all contributing databases should be freely accessible online.

The OMEGA metadatabase is therefore more than just a convenient way of locating and accessing other databases, it constitutes a major research tool in its own right, enabling the visualization and analysis of distributions on a global scale. At the time of writing, OMEGA includes metadata (approx. 49,000 records) from NODE, NANODe, DOAD and EANODe as well as other datasets representing southern Africa (supplied by Koen Martens), Australia (supplied by Chris Gouramanis) and China (supplied by Yangmin Qin). A partial dataset of c. 26,000 records compiled from NODE, DOAD and NANODe has been available for download since 2015, but data standards must be applied and very substantial taxonomic harmonization and validation of georeferenced localities remain to be completed before other data can be released (Horne, 2014).

6.6. Neotoma and North American Combined Ostracode Database

In 2009, a public access community-curated data resource became available, an international, collaborative database named Neotoma (named for packrats of the genus *Neotoma*) (Grimm et al., 2018; Williams et al., 2018). Neotoma (www.neotomadb.org) was developed to house, and access independently, cohesive paleoecological datasets of Pliocene through Modern age. The need for such an international resource with data stewards linked to databases and collections was driven by the growing problem, for many researchers, of maintaining funding to sustain and grow independent regional databases. A range of biological data for multi-proxy analysis can be found in Neotoma, including data access to pollen, vertebrates, diatoms and many other proxies for past environments and climates. These independent datasets retain their cohesiveness and are managed by data stewards associated with the datasets. The entire contents of NANODe and a large subset of DOAD are accessible in Neotoma. For DOAD, 4,053 georeferenced sites at which living ostracods were collected were ported into Neotoma, representing a subset of the 6,719 sites in DOAD (of which many recorded empty shells only). This allows a continental-scale biogeographic view of sites in which living ostracods paired with major ion hydrochemistry and limnologic data can be studied by examining the Canadian and U.S. datasets together. Additionally, numerous paleolimnologic ostracod records of Plio/Pleistocene and Holocene age are also housed in Neotoma. Amalgamated datasets from DOAD and NANODe, forming a North American Combined Ostracode Database (NACODe), were mapped and analyzed by Curry et al. (2012).

7. Paleobiology Database

The PBDB is an occurrence-based paleobiodiversity database, dedicated to fossil data of all time. PBDB collaborates closely with other important databases, e.g., GBIF (see Section 8), the Neotoma Database (see Section 6.6.), iDigBio, Macrostrat, Earth Life Consortium, and ePANDDA API. The freely downloadable data types from PBDB include georeferenced occurrences, specimens and measurements, geological strata, collections, diversity over time, taxa, opinions, and bibliographic references. PBDB relies on its authorized data enterers to enter data that mainly come from the published literature. For example, the top enterers for ostracod data entries are Matthew E. Clapham and John Alroy at the time of writing. Its fossil representation is partially influenced by what taxonomic groups the enterers focus on.

The PBDB has its own taxonomy system that is composed of authorities (i.e., references of taxonomic names combined with the authority) and opinions (i.e., references on the status of names and the relationships between names) (Peters and McClennen, 2016). An algorithm is used to produce the working taxonomy and to minimize influence from personal opinions (see the detailed algorithm in Peters and McClennen, 2016). The users could only influence the taxonomy by changing the rank of the opinion basis. Opinion references are ranked by their basis ("stated with evidence", "stated without evidence", "implied", and "second hand") and then by published year. The algorithm then selects the highest-ranked opinions for classification and creates the working taxonomy by using the information in the selected opinions: the reason of name structure or spelling (e.g., "original spelling", "recombination", "rank change", "misspelling") and name status (e.g., "belongs to", "synonym of", "nomen nudum"). Thereafter, occurrence data are dynamically linked to the constantly growing taxonomy database, so the taxonomy can be instantly updated. Database users can apply various filters that deal with open nomenclature, taxonomic reidentification, and updates on the accepted names. The taxonomic treatment is fully archived and downloadable. Together with WoRMS and about 100 taxonomic databases, PBDB also supplies its checklist dataset to GBIF (Paleobiology Database, 2021).

Collaborating with the GPlates software team (Müller et al., 2018), PBDB has a default plate motion model for calculating the paleocoordinates for data that have present-day coordinates and geological ages. The previous default model (until 2013) was provided by Christopher Scotese, and it is still available in the PBDB. Database users can toggle between Scotese's model and the current default model (since 2014), which is called the GPlates model (Wright et al., 2013) and is considered preferable in terms of accuracy.

PBDB has accumulated over 1.4 million occurrence records since its foundation in 1998. There are in total over 26,000 records of ostracods (Fig. 5A), the majority of which are marine ostracods. Ostracoda has fairly good number of occurrences with more than 100 genera in most geological periods but show decreasing numbers of data entries from the Paleozoic (63%) to the Mesozoic (20%) and the Cenozoic (17%) (Fig. 6). The most common levels of taxonomic resolution are genus (53%), species (39%), and class (5%). When cross checked with the WOD, about 20% of the accepted ostracod names from the PBDB are not currently registered in the WOD, and about 3% are marked as unaccepted in 2020. In the following analysis, we retain the classification in the PBDB.

Within Ostracoda, Podocopida, Palaeocopida, Platycopida, and Myodocopida have fossil records in the PBDB (Fig. 7). Podocopida and Palaeocopida have more fossil records than Platycopida, and Myodocopida. Palaeocopida is predominantly Paleozoic, whereas Podocopida increasingly dominate the fossil records towards the Cenozoic. These general trends are consistent with the chart presented in Fig. 20.14 in Armstrong and Brasier (2004) which was based on the data from Whatley et al. (1993).

8. GBIF

GBIF is an intergovernmental collaboration between over 100 official country/economy/organization participants (represented by the "nodes" in GBIF) that coordinate their own institutions and databases (called the "publishers") to share biodiversity data. The publishers share data through Integrated Publishing Toolkits which ensures the application of metadata standards, including Darwin Core and Ecological Metadata Language (see Introduction), on their datasets. Depending on the dataset nature, the publisher can choose from four dataset classes in

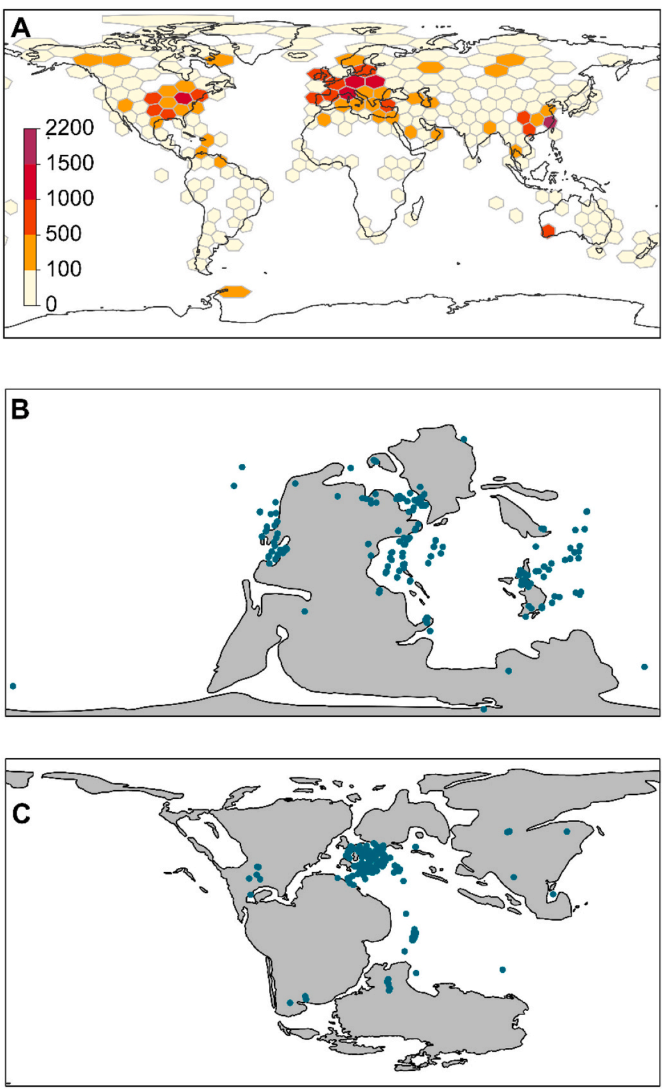

Fig. 5. Paleobiology Database (PBDB) ostracod records. (A) Geographical distribution of all records shown in a heatmap style. Total number of occurrence records: 21,111; number of locations: 2581. (B) Permian records in a paleogeographic map (275 Ma), (C) Jurassic records in a paleogeographic map (170 Ma). The maps are based on Scotese and Wright (2018).

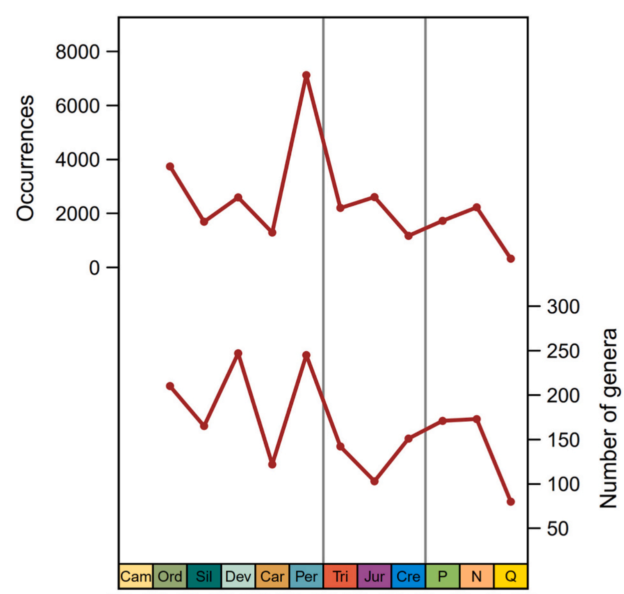

Fig. 6. Paleobiology Database: Number of ostracod occurrences and genera in the Phanerozoic. Abbreviations: Cambrian (Cam), Ordovician (Ord), Silurian (Sil), Devonian (Dev), Carboniferous (Car), Permian (Per), Triassic (Tri), Jurassic (Jur), Cretaceous (Cre), Paleogene (P), Neogene (N), Quaternary (Q), million years ago (Ma).

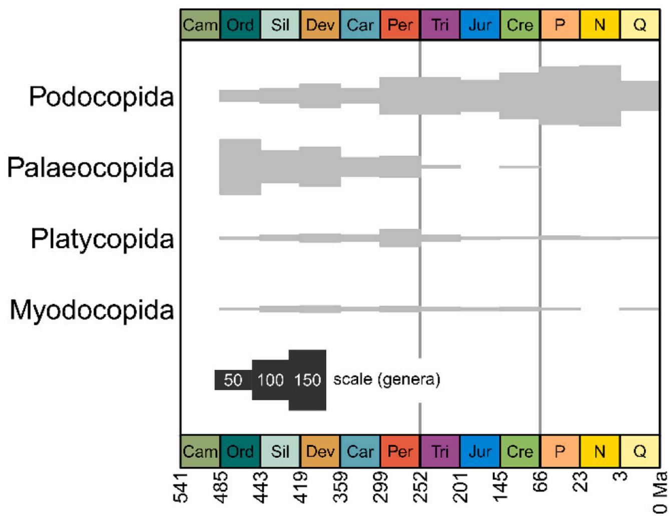

Fig. 7. Paleobiology Database: Number of genera per major orders of ostracods in the Phanerozoic. Abbreviations follow Fig. 6.

GBIF: metadata-only dataset for resources that have not been digitalized, taxonomic checklist, occurrence dataset, and sampling event dataset for data collected in greater details and with some protocol. Ingesting all kinds of data from data publishers, GBIF is the world's largest biodiversity database in terms of partnership and amount of data.

Both GBIF's taxonomy and occurrence systems aggregate data from its extensive network of nodes and data publishers. The GBIF Backbone Taxonomy is assembled based on about 100 taxonomic checklist sources (GBIF Secretariat, 2021). The top providers of taxonomic names include Catalogue of Life Checklist, International Barcode of Life project (iBOL) Barcode Index Numbers (BINs), Systema Dipterorum, The World Checklist of Vascular Plants (WCVP), WoRMS, and the PBDB. GBIF is designed for modern records and the majority of the GBIF occurrences are labeled as human observation. Many of the other occurrences are labeled as preserved specimens that come from botanical gardens and museums, e.g., departments of Botany, Entomology, Invertebrate Zoology and Vertebrate Zoology in the Smithsonian Institution. Although GBIF is not designed for paleobiological data, it has ingested occurrence datasets of fossil specimens from databases and museums, e. g., the Paleobiology Department in the Smithsonian Institution. These data are labeled as fossil specimens, but their geological ages have not been incorporated and their "year" column is blank in GBIF.

GBIF has over 376,000 occurrence records of ostracods at the time of writing. Preserved specimens account for ~43% of these records, indicating that museums are the biggest data sources of ostracods in GBIF. Fossil specimens account for ~36% of these records, and the top contributors are PANGAEA, PBDB, Smithsonian National Museum of Natural History, and Natural History Museum London. The rest are mainly labeled as human observation or material sample. Being the largest data aggregator, GBIF offers a convenient portal for checking what resources are available with which museums/institutions/databases and can potentially help ostracod researchers design new projects. On the other hand, users should be cautious about various issues that may result from aggregating a big number of datasets from different contexts.

9. Discussion: challenges and future directions

9.1. Taxonomic identification and harmonization

Regional datasets collected by one person or group may be internally consistent in terms of taxonomy and positional accuracy, but if they are to be utilized in a wider context in combination with other databases, then data standards must be established, and taxonomic harmonization is required. Geographically more extensive, literature-based datasets are likely to cover more of species' full distributions but, since they represent collections by many different people over many years, are more prone to inconsistencies in taxonomy as well as locational precision and accuracy. If publications include illustrations of the identified species, they can be checked and corrected where necessary, paying particular attention to synonyms and generic assignments; if unillustrated, assumptions must be made about the validity of the taxonomic names listed. The use of open nomenclature, often inconsistently, needs to be dealt with by either new identifications to species level or applying uncertainty filters in the database.

In the marine realm, AOD, OBIS, and PBDB display different approaches for taxonomic harmonization, corresponding to their different database scopes and architectures. Our investigation showed that AOD is one unique occurrence-based database that comes with high spatial coverage and fully harmonized taxonomy (albeit for a limited geographical region). This results from its focus on the Arctic ostracods and contribution from the same group of closely collaborated researchers. Taxonomic harmonization is more challenging for OBIS because it is a data harvester that recombines a vast amount of modern and Quaternary datasets from various data publishers at the global scale. Because occurrence data themselves contain no systematics information, OBIS uses WoRMS as its taxonomic backbone. PBDB also faces a more challenging task in taxonomic quality control, hosting fossil records at a longer geological time scale. It has an internal taxonomy database that uses an objective algorithm to combine authoritative names and taxonomic opinions.

For non-marine ostracod databases, the subjective global checklist of extant non-marine ostracod species published by Meisch et al. (2019) provides a standard against which all database records can be checked. Taxonomic harmonization efforts are currently focused on northern hemisphere (Holarctic) databases, supported at generic level by the recent publication of keys for the Nearctic (Smith and Horne, 2016) and Palaearctic (Horne et al., 2019) regions. Simple comparisons of species lists from DOAD, NANODe, NODE and EANODe suggest that at least 30 species are common to North America and Europe, and 10 of these are also common to East Asia. However, detailed investigations (in progress) show that while some are truly cosmopolitan, in many cases species listed under the same name turn out to be different species in different regions, while others are listed under different names that could be synonymized.

Another good solution is consulting an authoritative taxonomic database for taxonomic identification, taxonomic harmonization, and resolving outdated taxonomy. There is no silver bullet for taxonomic identification, and we will always need specialist expertise to conduct high standard taxonomic identification of ostracods. Nonetheless, a free, online database can strongly accelerate the traditionally slow work on taxonomic identification, and certainly coordinate updates in taxonomic opinions and consensus. Thanks to Kempf and many database contributors, valuable taxonomic resources have become available for current and future young scientists. The WOD is arguably the most promising platform for resolving the taxonomic challenges discussed above. WOD hosts freely available taxonomic information, descriptions, and illustrations for marine/non-marine and fossil/Recent ostracods. The collaboration between WoRMS and OBIS is an excellent example of how WOD can improve taxonomic quality in occurrence-based databases. Since all kinds of data in WOD are entered by experts, WOD needs more contributions from ostracodologists.

If a larger number of specialists invest only a few hours per week in WOD we would in one or a few years have as much freely available information as possible with copyright issues resolved on ostracod taxonomy and geographical and stratigraphical distribution. The specialists' contributions include (1) adding newly described species, registering new combinations and new classifications; (2) synonymies accompanied by references, uploading pdfs of publications related to ostracods; (3) scanning electron microscope photographs and other illustrations of each species, especially the type taxa. Another important feature missing in WOD is the publication of a higher (suprageneric) classification for ostracods by updating Whatley et al. (1993) and Horne et al. (2002). This was planned in the WOD editors' workshop in 2013, but still needs quite a lot of work. As differing opinions are common in taxonomy, all above tasks require careful coordination among the participating specialists to reach consensus opinions.

9.2. Validation of locality and sediment core section

Before they are made more widely available, databases that contain occurrence records should be validated by checking taxonomic and locality data for errors and making corrections where necessary. Most late 20th and 21st century publications provide adequate locality information, often including coordinates, while older literature frequently has shortcomings in this respect. Abstracting occurrence data from European literature for NODE, for example, has encountered many challenges including historical revision of political boundaries and place names, waterbodies from which ostracods were collected in the 19th century that were subsequently drained and no longer exist, and imprecise locality information (such as records from large lakes that do not specify a site within the lake, or small ponds only located by reference to the nearest town or village). Such issues can make compliance with data standards of integrative systems (e.g., OBIS, GBIF, Neotoma) very challenging. In preparation for making non-marine ostracod datasets available via the Neotoma and GBIF portals, current efforts are focused mainly on northern hemisphere databases (as with taxonomic harmonization), aided greatly by the growth of internet resources, particularly *Google Earth* and the online availability of historical maps (e.g., via *Digimap* in the UK); this work is time-consuming and for some records the confident assignment of precise coordinates may never be possible, but positional uncertainties of a few km may be considered acceptable when mapping distributions on large regional to global scales. A "traffic-light" system for indicating the validation status of species records, developed for NODE and OMEGA, might usefully be adopted by other database projects: green signifies accurate coordinates, amber denotes coordinates that are acceptable (as good as possible given limitations of the available data; comments may be added in justification, e.g., "coordinates approximate for center of lake"), and red indicates uncertain or unreliable records that need further checking.

In the case of ocean sediment cores, there are two critical problems related to the exact position of the samples. First, the sample ID may not be completely specified, especially in the older literature. For example, in order to update the age of a sample, it is important to specify the complete DSDP/ODP/IODP sample ID, which include the Site, Hole, Core, Section, and the position in the Section. There are typically several methods for calculating the depth below seafloor for each sample because the sediments expand after being recovered from the deep sea. The depth method used in an age model could be another different method because recombination of sections taken from multiple holes could be used to create a more complete sequence. Therefore, missing any of these ID elements or only specifying the depth below seafloor is losing the exact position of the samples. This may hinder the age assignment of the sample or comparison with other studies done on the same sediment cores. Second, OBIS may not recognize the paleo nature of the Quaternary downcore data in it. This could be an unintended consequence of the original design of OBIS. When the paleo data are combined with the true modern observation, the problem is clear: the year of observation might be labeled by the year of sampling cruise, but the time that the observation (i.e., the fossil assemblage) represents should be the geological age of the sample. Broadly similar problems are encountered with data from cored lake sediment archives and exposed sedimentary sequences on land. Such cases may require reconsideration of relevant data standards, leading to modification of existing ones or the introduction of new ones as the system evolves. Data standards are important, but they should not be allowed to exclude or confuse valuable data.

9.3. Paleolocality uncertainties

Fossil occurrences are georeferenced according to the present-day locations of the records, but these are not representative of their true distributions at the time when the fossils were living because tectonic plates have changed their position throughout the Phanerozoic. Uncertainty about the paleogeographic position of each occurrence is significant when mapping the distributions of fossil species. Paleodistributions can be mapped onto plate tectonic reconstructions (examples are shown in Fig. 5) but these are typically "snapshots" representing tens of millions of years so there are still significant uncertainties in the locations. While some regions have received extensive paleontological and paleogeographical investigation, and are likely to be well constrained (e.g., European region), others (e.g., central Asia) are less well constrained due to complex tectonics. For example, the paleogeographic position of the Uzbek myodocope ostracods occurrence of Mikhailova et al. (2020) during the late Silurian is debated. Thus, the authors tentatively regarded the Uzbek region/terrane in question as a small microcontinent placed between the North and South China plates based on fossil assemblages and paleomagnetic data (see Mikhailova et al., 2020). Additionally, there is very little record of Paleozoic deep-sea and open ocean planktonic faunas because the Paleozoic oceanic crust and the overlying sediment have been recycled via subduction since then (the oldest oceanic crust nowadays is less than 200 Myrs old). Thus, most of the deeper dwelling Paleozoic taxa will be preserved in shelf or slope sediments; deep basin records are generally very limited and restricted to tectonically complex settings such as nappes in accretion mountain ranges (e.g., Andes, Himalaya) and accretionary prisms around subduction zones (e.g., Japan) (Isozaki, 1997).

Another problem is uncertainty related to the age of a species and to the stratigraphic duration of its record. The taxa recognized in paleontology are skewed towards the more abundant, widespread, and geologically long-lived species, which have the greatest total number of individuals and occur in the greatest number of localities and rock types, and so are most likely to be preserved and recorded (Chaloner and Jablonski, 1994). Thus, longer-lived taxa will tend to be overrepresented in samples because there is a higher probability that at least one individual will be preserved (Solow and Smith, 1997).

9.4. Marine ostracod database and PBDB applications

Georeferenced occurrence databases are increasingly utilized in the field of macroecology and biogeography (e.g., Peters and McClennen, 2016; Williams et al., 2018). OBIS is a promising platform for global marine diversity research. However, our results showed that its ostracod coverage is concentrated in northern hemisphere and coastal areas. It currently provides little help in advancing our understanding of large-scale ecological patterns of ostracods. For example, latitudinal diversity gradients have only been recently explored in the North Atlantic, Arctic, and NW Pacific based on the AOD and individual datasets (Yasuhara et al., 2009; Yasuhara et al., 2012; Jöst et al., 2019; Chiu et al., 2020). Global analysis of ostracod macroecology and biogeography remains difficult. The latest global biogeographic scheme for marine ostracods was in the 1980s based on expert knowledge without quantitative data or analyses (Whatley, 1986; Titterton and Whatley, 1988). A much-needed updated global ostracod biogeography scheme would be greatly facilitated by global database developments (Yasuhara et al., 2019).

PBDB is the only occurrence database dedicated to integrating and curating deep-time fossil data. It is a resource for quantitatively estimating Phanerozoic-through ostracod diversity. Indeed, ostracods have a fair number of occurrences throughout the Ordovician to Present in PBDB (Fig. 6). Sepkoski (2000) was the first to utilize PBDB for estimating Crustacea *s.l.* diversity, recognizing ostracods as the single dominant component of crustacean fossil records throughout the Phanerozoic. Our generic diversity curve of ostracods using the up-to-date PBDB data is consistent with Sepkoski's (2000) in general

(Fig. 6). Phanerozoic generic diversity trends of ostracod orders reconstructed by using the up-to-date PBDB (Fig. 7) are also consistent with the ostracod experts' views (Whatley et al., 1993; Armstrong and Brasier, 2004). However, it is important to note that ostracods have received relatively little attention in the PBDB, and there is highly uneven distribution of data entries in space and time. Any estimations on the rates of origination and extinction of ostracods must be interpreted with caution.

PBDB is a promising platform for future collaboration in the ostracodology community. Many publications have explored macroevolutionary, paleoecological or biogeographical questions by analyzing georeferenced occurrence data, many of which have paleoenvironmental and trait information altogether in the PBDB (e.g., motility, primary skeletal mineralogy, life habitat, diet) (e.g., Marx and Uhen, 2010; Kiessling and Kocsis, 2015; Leprieur et al., 2016; Kocsis et al., 2018a; Reddin et al., 2020). In contrast, there have been very few ostracod studies utilizing PBDB (Donovan and van den Hoek Ostende, 2012; Forsey, 2016) even though Ostracoda is one of the few taxonomic groups that have abundant fossil records almost throughout the entire Phanerozoic (from Ordovician to Quaternary) (Figs. 6–7). The taxonomy scheme in the PBDB is designed to mitigate the issues of outdated taxonomy, open nomenclature, and conflicting opinions. Although the current taxonomic resources of ostracods are deficient, substantial future updates can presumably allow quantitative Phanerozoic-through biodiversity analyses.

9.5. Non-marine ostracod database applications

The combination of species occurrence records with geographical coordinates in non-marine distributional databases has high potential value as a research tool, for example in biogeographic studies of endemism and cosmopolitanism. This rich area of research holds many unanswered questions for which temporal and spatial biogeographic data are critical. Such datasets facilitate hypothesis-testing and can reveal interesting ecological as well as taxonomic insights. Calibrations of the temperature ranges of species, for use in Quaternary paleoclimatic reconstructions, are achieved by matching climate data to geographical occurrences in NODE, NANODe and DOAD (see, e.g., Horne et al., 2012; Marchegiano et al., 2020). Calibrations can be improved, moreover, by combining data from geographically separated regions in order to obtain more complete coverage of the climatic distributions of species. Molecular biological studies of geographical parthenogenesis in European non-marine ostracods have been supported by mapping sexual and asexual populations with the NODE database (e.g., Horne et al., 1998; Horne and Martens, 1998; Schmit et al., 2013).

10. Conclusions

In the rising trend of database research, it has become clear that occurrence and measurement data have more impact in science if they are digitalized, tabulated, and integrated. Even though they are far from comprehensive, ostracod data in many databases have demonstrated its unique potential in advancing our understanding of large-scale biodiversity and evolution in the future. They cover a wide range of ecosystems and are an important representative of Crustacea *s.l.* almost throughout the Phanerozoic in the fossil records. Therefore, we strongly recommend ostracodologists publish georeferenced data with suitable databases after they published their studies on taxonomy, ecology, biogeography, and evolution. Contributing ostracod data to databases can not only strongly impact paleontological and biodiversity research, but also add value to personal efforts and public funds.

Consistent and accurate taxonomy is of key importance to large-scale biodiversity research, while global databases are integrating an increasing number of datasets. Aphia's well-developed taxonomic system is the taxonomic backbone of many distributional databases. PBDB has an algorithm-powered taxonomic classification dynamically linked to its own occurrence data. Collaborating through Aphia and PBDB can be an effective way for ostracodologists to achieve the consensus opinions and directly improve the taxonomic quality of mainstream databases in the future.

CRediT authorship contribution statement

Huai-Hsuan M. Huang: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Moriaki Yasuhara:** Conceptualization, Writing – review & editing. **David J. Horne:** Conceptualization, Resources, Writing – review & editing. **Vincent Perrier:** Writing – review & editing. **Alison J. Smith:** Resources, Writing – review & editing. **Simone N. Brandão:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Robin J. Smith for the data of the East Asian Non-marine Ostracod Database; Renate Matzke-Karasz for the updated information on Kempf Database Ostracoda; Laura Gemery and Thomas Cronin for the updated information on Arctic Ostracode Database; two anonymous reviewers and two editors, Richard Jordan and Gene Hunt, for valuable comments. The NODE database was supported by Human Capital & Mobility Programme, European Union funding [grant number ERBCHRXCT930253]; Marie Curie Research and Training Network, European Union funding [grant number FP6-512492]; Contingency Fund award, European Union funding. NANODe and AJS were supported by the U. S. Geological Survey; Kent State University; National Science Foundation. AJS's work in Neotoma database was supported by National Science Foundation [grant numbers 0947459, 1550721, 1948297]; NSF EarthCube program [grant number 1540994]. HHMH and MY are partly supported by Research Grants Council, Hong Kong Special Administrative Region, China [grant numbers HKU 17300821, HKU 17300720, HKU 17302518, HKU 17311316]; the Seed Funding Programme for Basic Research, the University of Hong Kong [grant numbers 202011159122, 201811159076, 201711159057, 201611159053]; Faculty of Science RAE Improvement Fund, the University of Hong Kong; SKLMP Seed Collaborative Research Fund (grant number SKLMP/SCRF/0031), City University of Hong Kong; Seed Funding of the HKU-TCL Joint Research Centre for Artificial Intelligence. HHMH was supported by Ecology and Biodiversity Division Fund, the University of Hong Kong; Peter Buck Postdoc Fellowship, Smithsonian Institution. SNB's work on the World Ostracoda Database was/is funded by a Lifewatch Mini Grant and UESC [Projeto Biodiversidade de Ostracoda Marinho do Sul da Bahia].

References

Access to Biological Collections Data Task Group, 2005. Access to Biological Collection Data (ABCD). Biodiversity Information Standards (TDWG). Permanent IRI. http://www.tdwg.org/standards/115.

Alvarez Zarikian, C.A., 2015. Cenozoic bathyal and abyssal ostracods beneath the oligotrophic South Pacific Gyre (IODP Expedition 329 Sites U1367, U1368 and U1370). Palaeogeogr. Palaeoclimatol. Palaeoecol. 419, 115–142. https://doi.org/10.1016/j.palaeo.2014.07.024.

Angel, M., 2016. Personal library collection of Martin Angel of published and unpublished Halocyprid (Ostracoda) occurrences. Natural History Museum. Accessed on 2021-10-21 via https://obis.org/dataset/b1363fbb-afec-4dc7-85b6-39b8a5ebb55c.

Angel, M.V., Blachowiak-Samolyk, K., Chavtur, V.G., 2008. Atlas of Atlantic Planktonic Ostracods. Natural History Museum. Accessed on 2021-10-21 via. http://www.nhm.ac.uk/research-curation/scientific-resources/biodiversity/global-biodiversity/atlantic-ostracods/atlas/browse/index.jsp.

Armstrong, H.A., Brasier, M.D., 2004. Ostracods. Microfossils 219–248.
Balian, E., Segers, H., Lévêque, C., Martens, K., 2008. An introduction to the Freshwater Animal Diversity Assessment (FADA) project. Hydrobiologia 595, 3–8. https://doi.org/10.1007/s10750-007-9235-6.
Bergue, C., Govindan, A., 2010. Eocene-Pliocene deep sea ostracodes from ODP site 744A, Southern Indian Ocean. An. Acad. Bras. Cienc. 82, 747–760. https://doi.org/10.1590/S0001-37652010000300021.
biodiversity.aq, 2021. Antarctic Biodiversity Portal. Scientific Committee for Antarctic Research. Accessed on 2021-10-21 via www.biodiversity.aq.
Blachowiak-Samolyk, K., Angel, M., 2008. An Atlas of Southern Ocean planktonic ostracods. Southampton Oceanography Centre. Accessed on 2021-10-21 via http://deep.iopan.gda.pl/ostracoda/index.php.
Boomer, I., Horne, D.J., Slipper, I.J., 2003. The use of Ostracods in Palaeoenvironmental Studies, or what can you do with an Ostracod Shell? The Paleontological Society Papers 9, 153–180. https://doi.org/10.1017/S1089332600002199.
Brandão, S.N., 2012. Southern Ocean Ostracoda (except Halocypridina) Database. Accessed on 2021-10-22 via GBIF.org. https://doi.org/10.15468/de1pt3.
Brandão, S.N., Antonietto, L.S., Nery, D.G., Karanovic, I., 2020. World Ostracoda Database. https://doi.org/10.14284/364. Accessed on 2022-02-09 via https://www.marinespecies.org/ostracoda.
Brandt, A., Gooday, A.J., Brandão, S.N., Brix, S., Brökeland, W., Cedhagen, T., Choudhury, M., Cornelius, N., Danis, B., De Mesel, I., Diaz, R.J., Gillan, D.C., Ebbe, B., Howe, J.A., Janussen, D., Kaiser, S., Linse, K., Malyutina, M., Pawlowski, J., Raupach, M., Vanreusel, A., 2007. First insights into the biodiversity and biogeography of the Southern Ocean deep sea. Nature 447 (7142), 307–311. https://doi.org/10.1038/nature05827.
Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., Schouppe, M., 2019. Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. Frontiers in Big Data 2, 43. https://doi.org/10.3389/fdata.2019.00043.
Chaloner, W.G., Jablonski, D., 1994. Extinctions in the Fossil Record: Discussion. Philosophical Transactions of the Royal Society of London Series B 344 (1307), 16–17.
Chaudhary, C., Richardson, A.J., Schoeman, D.S., Costello, M.J., 2021. Global warming is causing a more pronounced dip in marine species richness around the equator. PNAS 118 (15), e2015094118. https://doi.org/10.1073/pnas.2015094118.
Chiu, W.-T.R., Yasuhara, M., Cronin, T.M., Hunt, G., Gemery, L., Wei, C.-L., 2020. Marine latitudinal diversity gradients, niche conservatism and out of the tropics and Arctic: Climatic sensitivity of small organisms. J. Biogeogr. 47 (4), 817–828. https://doi.org/10.1111/jbi.13793.
Costello, M.J., 2000. Developing species information systems: the European Register of Marine Species (ERMS). Oceanography 13 (3), 48–55.
Costello, M.J., Wieczorek, J., 2014. Best practice for biodiversity data management and publication. Biol. Conserv. 173, 68–73. https://doi.org/10.1016/j.biocon.2013.10.018.
Costello, M.J., Emblow, C., White, R.J., 2001. European Register of Marine Species: A Check-List of the Marine Species in Europe and a Bibliography of Guides to their Identification. Muséum national d'histoire naturelle, Paris.
Cronin, T.M., Briggs Jr., W., Brouwers, E., Whatley, R., Wood, A., Cotton, M., 1991. Modern Arctic podocopid ostracode database. In: US Geological Survey Open-File Report 91–385, pp. 1–51.
Cronin, T.M., Holtz, T.R., Brouwers, E.M., Briggs, W.M., Whatley, R.C., Wood, A., 1995. Arctic Ostracode Database. In: U.S. Geological Survey, Reston, VA. Accessed on 2021-10-21 via https://geo-nsdi.er.usgs.gov/metadata/other/prism-arctic-ostracod/.
Cronin, T.M., Gemery, L., Briggs, W.M., Jakobsson, M., Polyak, L., Brouwers, E.M., 2010. Quaternary Sea-ice history in the Arctic Ocean based on a new Ostracode Sea-ice proxy. Quat. Sci. Rev. 29 (25–26), 3415–3429. https://doi.org/10.1016/j.quascirev.2010.05.024.
Cronin, T.M., Gemery, L.J., Briggs, W.M., Brouwers, E.M., Schornikov, E.I., Stepanova, A., Wood, A.M., Yasuhara, M., Siu, S., 2021. Arctic Ostracode Database 2020. NOAA's National Centers for Environmental Information (NCEI). Accessed on 2021-10-14 via https://www.ncdc.noaa.gov/paleo/study/32312.
Curry, B.B., Delorme, L.D., Smith, A.J., Palmer, D.F., Stiff, B.J., 2012. The Biogeography and Physicochemical Characteristics of Aquatic Habitats of Freshwater Ostracods in Canada and the United States. In: Horne, D.J., Holmes, J.A., Rodriguez-Lazaro, J., Viehberg, F.A. (Eds.), Ostracoda as Proxies for Quaternary Climate Change. Elsevier, Developments in Quaternary Sciences, pp. 85–115.
Cuvelier, D., Claus, S., Appeltans, W., Vanhoorne, B., Vanden Berghe, E., Costello, M., 2006. European Register of Marine Species (ERMS)-plans turning into reality. MarBEF Newsletter 4, 14–15.
Darwin Core Task Group, 2009. Darwin Core. Permanent IRI, Biodiversity Information Standards (TDWG). http://www.tdwg.org/standards/450.
Davies, T.G., Rahman, I.A., Lautenschlager, S., Cunningham, J.A., Asher, R.J., Barrett, P. M., Bates, K.T., Bengtson, S., Benson, R.B., Boyer, D.M., 2017. Open data and digital morphology. Proc. R. Soc. B Biol. Sci. 284 (1852), 20170194. https://doi.org/10.1098/rspb.2017.0194.
Donovan, S.K., van den Hoek Ostende, L.W., 2012. *Neoextrania*, a replacement name for *Extrania* Sun & Wang, 1985 (Ostracoda), preoccupied by *Extrania* Qian in Qiu et al., 1983 (Trilobita). Alcheringa: An Australasian Journal of Palaeontology 36 (3), 423. https://doi.org/10.1080/03115518.2012.694304.
Ellis, B.F., Messina, A.R., 1952. Catalogue of Ostracoda. American Museum of Natural History Special Publications, New York.
Faith, D., Collen, B., Ariño, A., Koleff, P.K.P., Guinotte, J., Kerr, J., Chavan, V., 2013. Bridging the biodiversity data gaps: Recommendations to meet users' data needs. Biodivers. Inform. 8 (2) https://doi.org/10.17161/bi.v8i2.4126.
Fegraus, E.H., Andelman, S., Jones, M.B., Schildhauer, M., 2005. Maximizing the value of ecological data with structured metadata: an introduction to ecological metadata language (EML) and principles for metadata creation. Bull. Ecol. Soc. Am. 86 (3), 158–168.
Forester, R.M., Smith, A.J., Palmer, D.F., Curry, B.B., 2005. North American Nonmarine Ostracode Database, Version 1, www.kent.edu/NANODe. Kent State University, Kent, Ohio, U.S.A.
Forsey, G., 2016. Ostracods as proxies for past seagrass: a review. Palaeogeogr. Palaeoclimatol. Palaeoecol. 447, 22–28. https://doi.org/10.1016/j.palaeo.2016.01.028.
GBIF, 2021. What is GBIF? Accessed on 2021-10-14 via https://www.gbif.org/what-is-gbif.
GBIF Secretariat, 2021. GBIF Backbone Taxonomy. Checklist dataset. https://doi.org/10.15468/39omei. Accessed on 2021-10-21 via GBIF.org.
GBIF/TDWG Multimedia Resources Task Group, 2013. Audubon Core Multimedia Resources Metadata Schema. Permanent IRI, Biodiversity Information Standards (TDWG). http://www.tdwg.org/standards/638.
Gemery, L., Cronin, T.M., Briggs, W.M., Brouwers, E.M., Schornikov, E.I., Stepanova, A., Wood, A.M., Yasuhara, M., 2017. An Arctic and Subarctic ostracode database: biogeographic and paleoceanographic applications. Hydrobiologia 786 (1), 59–95.
Gemery, L., Cooper, L.W., Magen, C., Cronin, T.M., Grebmeier, J.M., 2021a. Stable oxygen isotopes in shallow marine ostracodes from the northern Bering and Chukchi Seas. Mar. Micropaleontol., 102001 https://doi.org/10.1016/j.marmicro.2021.101979.
Gemery, L., Cronin, T.M., Cooper, L.W., Dowsett, H.J., Grebmeier, J.M., 2021b. Biogeography and ecology of Ostracoda in the US northern Bering, Chukchi, and Beaufort Seas. PLoS One 16 (5), e0251164. https://doi.org/10.1371/journal.pone.0251164.
Goddard, A., Wilson, N., Cryer, P., Yamashita, G., 2011. Data hosting infrastructure for primary biodiversity data. BMC bioinformatics 12 (S15), S5.
Grimm, E.C., Blois, J., Giesecke, T., Graham, R., Smith, A., Williams, J., 2018. Constituent databases and data stewards in the Neotoma Paleoecology Database: history, growth, and new directions. PAGES Magazine 26, 64–65.
Hendricks, J.R., Stigall, A.L., Lieberman, B.S., 2015. The Digital Atlas of Ancient Life: Delivering Information on Paleontology and Biogeography. Palaeontol. Electron. 1–9 https://doi.org/10.26879/153E.
Holetschek, J., Dröge, G., Güntsch, A., Berendsohn, W., 2012. The ABCD of primary biodiversity data access. Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology 146 (4), 771–779. https://doi.org/10.1080/11263504.2012.740085.
Holmes, J.A., Chivas, A.R., 2002. Introduction. In: Holmes, J.A., Chivas, A.R. (Eds.), The Ostracoda: Applications in Quaternary Research. American Geophysical Union, pp. 1–4.
Horne, D.J., 2014. Ostracod Metadatabase of Environmental and Geographical Attributes (Europe, North America). BioFresh. Accessed on 2021-11-05 via http://data.freshwaterbiodiversity.eu. https://doi.org/10.13148/bfcf15.
Horne, D.J., Martens, K., 1998. Geographical parthenogenesis in European non-marine ostracods: Post-glacial invasion or Holocene stability? Hydrobiologia 391, 1–7. https://doi.org/10.1023/A:1003508210166.
Horne, D., Baltanas, A., Paris, G., 1998. Geographical distribution of reproductive modes in living non-marine ostracods. In: Martens, K. (Ed.), Sex and Parthenogenesis: Evolutionary Ecology of Reproductive Modes in Non-marine Ostracods. Backhuys, Leiden, pp. 77–99.
Horne, D.J., Cohen, A., Martens, K., 2002. Taxonomy, morphology and biology of Quaternary and living Ostracoda. In: Holmes, J.A., Chivas, A.R. (Eds.), The Ostracoda: Applications in Quaternary Research. American Geophysical Union, Washington, DC, pp. 5–36.
Horne, D.J., Curry, B.B., Delorme, L.D., Martens, K., Smith, A.J., Smith, R.J., 2011. OMEGA: the Ostracod metadatabase of environmental and geographical attributes. Joannea Geologie und Paläontologie 11, 80–84.
Horne, D.J., Curry, B.B., Mesquita-Joanes, F., 2012. Mutual Climatic Range Methods for Quaternary Ostracods. In: Horne, D.J., Holmes, J.A., Rodriguez-Lazaro, J., Viehberg, F.A. (Eds.), Developments in Quaternary Sciences. Elsevier, pp. 65–84.
Horne, D.J., Meisch, C., Martens, K., 2019. Arthropoda: Ostracoda. In: Rogers, D.C., Thorp, J. (Eds.), Thorp and Covich's Freshwater Invertebrates, Fourth ed. Academic Press, pp. 725–760.
Isozaki, Y., 1997. Permo-Triassic boundary superanoxia and stratified superocean: records from lost deep sea. Science 276 (5310), 235–238.
Jablonski, D., Sepkoski Jr., J.J., 1996. Paleobiology, community ecology, and scales of ecological pattern. Ecology 77 (5), 1367–1378.
Jones, M.B., O'Brien, M., Mecum, B., Boettiger, C., Schildhauer, M., Maier, M., Whiteaker, T., Earl, S., Chong, S., 2019. Ecological metadata language version 2.2. 0. KNB Data Repository. https://doi.org/10.5063/F11834T2.
Jöst, A.B., Yasuhara, M., Wei, C.L., Okahashi, H., Ostmann, A., Martínez Arbizu, P., Mamo, B., Svavarsson, J., Brix, S., 2019. North Atlantic Gateway: Test bed of deep-sea macroecological patterns. J. Biogeogr. 46 (9), 2056–2066. https://doi.org/10.1111/jbi.13632.
Kiessling, W., Kocsis, Á.T., 2015. Biodiversity dynamics and environmental occupancy of fossil azooxanthellate and zooxanthellate scleractinian corals. Paleobiology 41 (3), 402–414. https://doi.org/10.1017/pab.2015.6.
Kocsis, Á.T., Reddin, C.J., Kiessling, W., 2018a. The biogeographical imprint of mass extinctions. Proc. R. Soc. B Biol. Sci. 285 (1878), 20180232. https://doi.org/10.1098/rspb.2018.0232.
Kocsis, Á.T., Reddin, C.J., Kiessling, W., 2018b. The stability of coastal benthic biogeography over the last 10 million years. Glob. Ecol. Biogeogr. 27 (9), 1106–1120. https://doi.org/10.1111/geb.12771.
Lautenschlager, S., Rücklin, M., 2014. Beyond the print—virtual paleontology in science publishing, outreach, and education. J. Paleontol. 88 (4), 727. https://doi.org/10.1666/13-085.

Lazarus, D., 1994. Neptune: a marine micropaleontology database. Math. Geol. 26 (7), 817–832.

Lazarus, D., Barron, J., Renaudie, J., Diver, P., Türke, A., 2014. Cenozoic planktonic marine diatom diversity and correlation to climate change. PLoS One 9 (1), e84857. https://doi.org/10.1371/journal.pone.0084857.

Leprieur, F., Descombes, P., Gaboriau, T., Cowman, P.F., Parravicini, V., Kulbicki, M., Melián, C.J., De Santana, C.N., Heine, C., Mouillot, D., Bellwood, D.R., Pellissier, L., 2016. Plate tectonics drive tropical reef biodiversity dynamics. Nat. Commun. 7 (1), 11461. https://doi.org/10.1038/ncomms11461.

Liow, L.H., Stenseth, N.C., 2007. The rise and fall of species: implications for macroevolutionary and macroecological studies. Proc. R. Soc. B Biol. Sci. 274 (1626), 2745–2752. https://doi.org/10.1098/rspb.2007.1006.

Lockwood, R., Cohen, P.A., Uhen, M.D., Ryker, K., 2018. Utilizing the Paleobiology Database to Provide Educational Opportunities for Undergraduates. Cambridge University Press.

Majoran, S., Dingle, R.V., 2001. Cenozoic deep-sea ostracods from southwestern South Atlantic (DSDP/ODP Sites 329, 513 and 699). Revista Espanola de Micropaleontologia 33 (2), 205–216.

Marchegiano, M., Horne, D.J., Gliozzi, E., Francke, A., Wagner, B., Ariztegui, D., 2020. Rapid late Pleistocene climate change reconstructed from a lacustrine ostracod record in Central Italy (Lake Trasimeno, Umbria). Boreas 49 (4), 739–750. https://doi.org/10.1111/bor.12450.

Martens, K., Schön, I., Meisch, C., Horne, D.J., 2008. Global diversity of ostracods (Ostracoda, Crustacea) in freshwater. Hydrobiologia 595 (1), 185–193. https://doi.org/10.1007/s10750-007-9245-4.

Martens, K., Savatenalinton, S., Schön, I., Meisch, C., Horne, D.J., 2013. World Checklist of Freshwater Ostracoda Species. World Wide Web Electronic Publication. Accessed on 2021-11-05 via http://fada.biodiversity.be/group/show/18.

Martins, M.J.F., Hunt, G., Thompson, C.M., Lockwood, R., Swaddle, J.P., Puckett, T.M., 2020. Shifts in sexual dimorphism across a mass extinction in ostracods: implications for sexual selection as a factor in extinction risk. Proc. R. Soc. B 287 (1933), 20200730. https://doi.org/10.1098/rspb.2020.0730.

Marx, F.G., Uhen, M.D., 2010. Climate, Critters, and Cetaceans: Cenozoic Drivers of the Evolution of Modern Whales. Science 327 (5968), 993–996. https://doi.org/10.1126/science.1185581.

Matzke-Karasz, R., 2014. Eugen Karl Kempf, the man behind the Kempf Database Ostracoda. Crustaceana 87 (8–9), 901–922.

MarBEF, 2004. European Marine Biodiversity Datasets. Accessed on 2022-02-10 via http://www.marbef.org/data/imis.php?module=dataset.

Meisch, C., Smith, R.J., Martens, K., 2019. A subjective global checklist of the extant non-marine Ostracoda (Crustacea). European Journal of Taxonomy 492. https://doi.org/10.5852/ejt.2019.492.

Mesquita-Joanes, F., Smith, A.J., Viehberg, F.A., 2012. Chapter 2 - the ecology of ostracoda across levels of biological organisation from individual to ecosystem: a review of recent developments and Future potential. In: Horne, D.J., Holmes, J.A., Rodriguez-Lazaro, J., Viehberg, F.A. (Eds.), Developments in Quaternary Sciences. Elsevier, pp. 15–35.

Micropaleontology Press, 2021. Ellis and Messina Catalogues. Accessed on 2021-11-05 via https://www.micropress.org/ellis-messina.html.

Mikhailova, E., Perrier, V., Williams, M., Siveter, D.J., Tarasenko, A., Salimova, F., Kim, I.A., 2020. Cosmopolitan myodocope ostracods from the Silurian of Uzbekistan, Central Asia. Les ostracodes myodocopes cosmopolites du silurien Asie Centrale. Bulletin de la Société Géologique de France 191 (1), 32.

Morris, R.A., Barve, V., Carausu, M., Chavan, V., Cuadra, J., Freeland, C., Hagedorn, G., Leary, P., Mozzherin, D., Olson, A., 2013. Discovery and publishing of primary biodiversity data associated with multimedia resources: the Audubon Core strategies and approaches. Biodivers. Inform. 8 (2) https://doi.org/10.17161/bi.v8i2.4117.

Müller, R.D., Cannon, J., Qin, X., Watson, R.J., Gurnis, M., Williams, S., Pfaffelmoser, T., Seton, M., Russell, S.H., Zahirovic, S., 2018. GPlates: building a virtual Earth through deep time. Geochem. Geophys. Geosyst. 19 (7), 2243–2261. https://doi.org/10.1029/2018GC007584.

Nelson, G., Ellis, S., 2019. The history and impact of digitization and digital data mobilization on biodiversity research. Philos. Trans. R. Soc. B 374 (1763), 20170391. https://doi.org/10.1098/rstb.2017.0391.

OBIS, 2021. Ocean Biodiversity Information System. Intergovernmental Oceanographic Commission of UNESCO. Accessed on 2021-11-05 via https://obis.org/.

OBIS Secretariat, 2021. OBIS Secretariat. Ocean Biogeographic Information System. Accessed on 2021-10-14 via https://obis.org/node/310922b4-9d0c-4de1-92d7-9b442d34765b.

Paleobiology Database, 2021. The Paleobiology Database. Checklist Dataset. Accessed on 2021-10-21 via GBIF.org. https://doi.org/10.15468/zzoyxi.

PANGAEA®, 2021. Data Publisher for Earth & Environmental Science. https://doi.org/10.1594/PANGAEA.

Perrier, V., Williams, M., Siveter, D.J., 2015. The fossil record and palaeoenvironmental significance of marine arthropod zooplankton. Earth Sci. Rev. 146, 146–162. https://doi.org/10.1016/j.earscirev.2015.02.003.

Peters, S.E., McClennen, M., 2016. The Paleobiology Database application programming interface. Paleobiology 42 (1), 1–7. https://doi.org/10.1017/pab.2015.39.

Powers, S.M., Hampton, S.E., 2019. Open science, reproducibility, and transparency in ecology. Ecol. Appl. 29 (1), e01822. https://doi.org/10.1002/eap.1822.

Reddin, C.J., Kocsis, Á.T., Aberhan, M., Kiessling, W., 2020. Victims of ancient hyperthermal events herald the fates of marine clades and traits under global warming. Glob. Chang. Biol. 27 (4), 868–878. https://doi.org/10.1111/gcb.15434.

Rick, T.C., Lockwood, R., 2013. Integrating paleobiology, archeology, and history to inform biological conservation. Conserv. Biol. 27 (1), 45–54. https://doi.org/10.1111/j.1523-1739.2012.01920.x.

Robertson, T., Döring, M., Guralnick, R., Bloom, D., Wieczorek, J., Braak, K., Otegui, J., Russell, L., Desmet, P., 2014. The GBIF integrated publishing toolkit: facilitating the efficient publishing of biodiversity data on the internet. PLoS One 9 (8), e102623. https://doi.org/10.1371/journal.pone.0102623.

Rodriguez-Lazaro, J., Ruiz-Muñoz, F., 2012. chapter 1 - a general introduction to ostracods: morphology, distribution, fossil record and applications. In: Horne, D.J., Holmes, J.A., Rodriguez-Lazaro, J., Viehberg, F.A. (Eds.), Developments in Quaternary Sciences. Elsevier, pp. 1–14.

Schmit, O., Bode, S.N.S., Camacho, A., Horne, D.J., Lamatsch, D.K., Martens, K., Martins, M.J.F., Namiotko, T., Rossetti, G., Rueda-Sevilla, J., Schön, I., Vandekerkhove, J., Mesquita-Joanes, F., 2013. Linking present environment and the segregation of reproductive modes (geographical parthenogenesis) in *Eucypris virens* (Crustacea: Ostracoda). J. Biogeogr. 40 (12), 2396–2408. https://doi.org/10.1111/jbi.12174.

Scholes, R., Mace, G., Turner, W., Geller, G., Jürgens, N., Larigauderie, A., Muchoney, D., Walther, B., Mooney, H., 2008. Toward a global biodiversity observing system. Science 321 (5892), 1044–1045. https://doi.org/10.1126/science.1162055.

Scotese, C.R., Wright, N., 2018. PALEOMAP Paleodigital Elevation Models (PaleoDEMS) for the Phanerozoic. https://www.earthbyte.org/paleodem-resource-scotese-and-wright-2018.

Sepkoski, J.J.Jr., 2000. Crustacean biodiversity through the marine fossil record. Contributions to Zoology 69, 213–222.

Smith, A.J., Horne, D.J., 2002. Ecology of marine, marginal marine and nonmarine ostracodes. Geophysical Monograph Series 37–64.

Smith, A.J., Horne, D.J., 2016. Class Ostracoda. In: Thorp, J., Rogers, D.C. (Eds.), Thorp and Covich's Freshwater Invertebrates, Fourth ed. Academic Press, pp. 477–514.

Smith, A.J., Horne, D.J., Martens, K., Schön, I., 2015. Class Ostracoda. In: Thorp, J.H., Rogers, D.C. (Eds.), Thorp and Covich's Freshwater Invertebrates, Fourth ed. Academic Press, Boston, pp. 757–780.

Solow, A.R., Smith, W., 1997. On fossil preservation and the stratigraphic ranges of taxa. Paleobiology 271–277. https://doi.org/10.1017/S0094837300019680.

Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V., Worm, B., 2010. Global patterns and predictors of marine biodiversity across taxa. Nature 466, 1098–1101. https://doi.org/10.1038/nature09329.

Titterton, R., Whatley, R.C., 1988. The Provincial Distribution of Shallow Water Indo-Pacific Marine Ostracoda: Origins, Antiquity, Dispersal Routes and Mechanisms, Developments in Palaeontology and Stratigraphy, pp. 759–786.

Vandepitte, L., Vanhoorne, B., Decock, W., Dekeyzer, S., Trias Verbeeck, A., Bovit, L., Hernandez, F., Mees, J., 2015. How Aphia—the platform behind several online and taxonomically oriented databases—can serve both the taxonomic community and the field of biodiversity informatics. Journal of Marine Science and Engineering 3 (4), 1448–1473. https://doi.org/10.3390/jmse3041448.

Viehberg, F.A., Matzke-Karasz, R., Boush, L.P., Smith, A.J., 2014. Fossil and recent meet Kempf database. Crustaceana 87 (8–9), 897–900.

Vilhena, D.A., Antonelli, A., 2015. A network approach for identifying and delimiting biogeographical regions. Nat. Commun. 6 (1), 6848. https://doi.org/10.1038/ncomms7848.

Villéger, S., Blanchet, S., Beauchard, O., Oberdorff, T., Brosse, S., 2011. Homogenization patterns of the world's freshwater fish faunas. PNAS 108, 18003–18008. https://doi.org/10.1073/pnas.1107614108.

Weibel, S., 1999. The State of the Dublin Core Metadata Initiative: April 1999. Bull. Am. Soc. Inf. Sci. Technol. 25 (5), 18–22. https://doi.org/10.1002/bult.127.

Whatley, R., 1986. The southern end of Tethys: An important locus for the origin and evolution of both deep and shallow water Ostracoda. In: McKenzie, K.G. (Ed.), Proceedings 2nd Shallow Tethys Symposium. Wagga Wagga, Bolkema, pp. 461–475.

Whatley, R.C., Siveter, D.J., Boomer, I.D., 1993. Arthropoda (Crustacea: Ostracoda). In: Benton, M.J. (Ed.), The Fossil Record 2. Chapman & Hall, London, pp. 343–356.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., Vieglais, D., 2012. Darwin Core: an evolving community-developed biodiversity data standard. PLoS One 7 (1), e29715. https://doi.org/10.1371/journal.pone.0029715.

Williams, M., Siveter, D.J., Salas, M.J., Vannier, J., Popov, L.E., Pour, M.G., 2008. The earliest ostracods: the geological evidence. Senckenb. Lethaea 88 (1), 11–21.

Williams, J.W., Grimm, E.C., Blois, J.L., Charles, D.F., Davis, E.B., Goring, S.J., Graham, R.W., Smith, A.J., Anderson, M., Arroyo-Cabrales, J., 2018. The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource. Quat. Res. 89 (1), 156–177. https://doi.org/10.1017/qua.2017.105.

Wright, N., Zahirovic, S., Müller, R., Seton, M., 2013. Towards community-driven paleogeographic reconstructions: integrating open-access paleogeographic and paleobiology data with plate tectonics. Biogeosciences 10 (3), 1529–1541. https://doi.org/10.5194/bg-10-1529-2013.

Yasuhara, M., Hunt, G., Cronin, T.M., Okahashi, H., 2009. Temporal latitudinal-gradient dynamics and tropical instability of deep-sea species diversity. PNAS 106 (51), 21717–21720. https://doi.org/10.1073/pnas.0910935106.

Yasuhara, M., Hunt, G., van Dijken, G., Arrigo, K.R., Cronin, T.M., Wollenburg, J.E., 2012. Patterns and controlling factors of species diversity in the Arctic Ocean. J. Biogeogr. 39 (11), 2081–2088. https://doi.org/10.1111/j.1365-2699.2012.02758.x.

Yasuhara, M., 2019. Marine biodiversity in space and time: what tiny fossils tell. Mètode Science Studies Journal-Annual Review 9, 77–81.

Yasuhara, M., Cronin, T.M., 2008. Climatic influences on deep-sea ostracode (Crustacea) diversity for the last three million years. Ecology 89 (sp11), S53–S65. https://doi.org/10.1890/07-1021.1.

Yasuhara, M., Hong, Y., Tian, S.Y., Chong, W.K., Okahashi, H., Littler, K., Cotton, L., 2019. Eocene shallow-marine ostracods from Madagascar: southern end of the Tethys? J. Syst. Palaeontol. 17 (9), 705–757. https://doi.org/10.1080/14772019.2018.1453555.